THE BRITISH JOURNAL OF PSYCHOLOGY

Statistical Section

EDITED BY
CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ARTHUR SUMMERFIELD
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
B. BABINGTON SMITH
W. STEPHENSON
P. E. VERNON

Managing Editor : H. G. MAULE

Volume III
1950

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4.

Digitized by the Internet Archive
in 2026

https://archive.org/details/british-journal-of-psychology_1950-03_3_part-1

AN APPLICATION OF FACTORIAL ANALYSIS TO THE STUDY OF TEST ITEMS

By P. E. VERNON
Institute of Education, University of London

I. *Introduction.* II. *Data and Methods.* III. *Analysis of Items.* IV. *Analysis of Tests.* V. *Functional Content of Item Factors.* VI. *Functional Content of Separate Items.* VII. *Final Analysis of Formal Factors.* VIII. *Conclusions Regarding Factor Analysis.* IX. *Summary.*

I. INTRODUCTION

In an earlier article (Vernon, 25), it was suggested that factor analysis could be useful in analysing the homogeneity of content of test items, though the amount of work it involves limits its application to short tests. The present study was designed to discover how valuable would be the information obtainable about the items of a test of moderate length, namely, Progressive Matrices. This test was chosen, not with any intent to criticize, but because :

(*a*) it is so widely used that fuller understanding of what it measures is desirable ;

(*b*) answer sheets for the 20-minute version from 640 recruits, together with scores on 21 other tests, happened to be available ;

(*c*) the content of the test appears, from inspection, to be, if anything, more homogeneous than that of many other popular intelligence, attainment, and aptitude tests. Hence, if appreciable heterogeneity is found, we might need to reconsider the soundness of current methods of test construction.

Several researches have in fact indicated that the items of which certain tests are composed do measure diverse abilities. Burt (1), Wright (29), Burt and John (3), and McNemar (12) have shown that the Stanford-Binet or Terman-Merrill scales embody several group factors over and above their general content. In McNemar's 14 analyses at different age levels the mean first factor variance was close to 40 per cent. and the mean variance of his bipolars 13 per cent. But the latter were exaggerated by unreliable correlations from small groups, and amounted only to 10·5 per cent. in analyses where 200 cases were available.

The relative sizes of general and bipolar variances naturally depend on the heterogeneity of the subjects tested. But one would suggest that, in a representative group, the former should usually exceed the latter by at least four times. That this is a moderate demand is shown by Eysenck and Crown's (5) analysis of a 24-item attitude scale. There the general and bipolar factor variances were 48 per cent. and 5 per cent. Another research on the Stanford-Binet test, by Wright, yielded variances of 60·8 per cent. and 17·3 per cent. among ten-year olds. But as the last two of Wright's six bipolars were almost certainly non-significant, it would be fairer to take account only of the 12·9 per cent. variance of the four chief ones. The same criterion may be applied to the sub-tests of a battery. For example, the Army Summed S.G. is based on scores from five tests. When analysed in a representative group of recruits the first factor variance of these tests is 66·6, and three bipolars contribute 10·4 per cent. Had the ratio only reached, say, 3 to 1, we should be much more doubtful about employing a summed score.

An example of heterogeneity is provided by Guilford's (8) analysis of the ten sets of items in the Seashore Pitch Discrimination test, given to 300 students. He concludes that the more difficult sets (5 to 0·5 cycles) depend on entirely different factors from the easier sets (30 to 8 cycles). In spite of the test's apparently homogeneous material, it does not measure one and the same pitch ability throughout. An alternative possibility is that more able subjects use very different discriminatory methods from the less able. However Guilford notes that the two most difficult sets (1 and 0·5 cycles) are highly unreliable ; and if these are omitted the first factor variance of the remainder is 57·8 per cent., bipolar variance 15·8 per cent. Indeed, with a less selected group (for Pitch Discrimination correlates appreciably with intelligence), the general factor in these sets might easily have satisfied our suggested criterion.

There is other evidence of ' difficulty factors.' Burt and John, in their analysis of Stanford-Binet X- and XII-yr. items, obtained a factor separating almost all the former from the latter items. Hertzman (10) studied correlations between the easier and more difficult halves of a battery of group-tests, and concluded that they measure somewhat different abilities. Ferguson (6) points out that spurious factors corresponding to the difficulty of different sets of items will be introduced if Pearsonian product-moment or point inter-correlations are analysed. Wherry and Gaylord (28) recommend substituting tetrachoric correlations, which are unaffected by item-difficulty ; but in point of fact this had already been done both by Burt and Guilford in the researches just mentioned.

Another possibility is that the form of the items (open-type, analogies, classification, etc.) introduces unwanted factors. Guilford (7) claims that this is unimportant, and cites Smith's (18) investigation, where 14 verbal, numerical, and spatial tests in analogies, classification, and open-response form were compared. Although Smith's application of tetrad analysis failed to reveal any clear-cut formal factors, a re-analysis of his figures by the present writer, using group-factor technique, suggests the presence of a group factor with some 7 per cent. variance among multiple-choice tests (analogies *and* classifications) not present in open-response tests. The total variance of the factors based on the content of the tests among college students approximated 50 per cent. Most educational psychologists seem to regard an individual Stanford or Terman-Merrill I.Q. as having greater practical utility than an I.Q. from an equally reliable group-test, and this may be because the latter is distorted by some such formal factor, i.e., by the ability to do multiple-choice tests, which has been described by the writer elsewhere as ' test sophistication ' (Vernon, 23).

Thurstone (22) and Shaefer (15) have suggested that the perceptual speed factor, *P*, appears only among tests composed of easy items. Though it was present in tests of matching or identifying printed letters and numbers, diagrams and pictures among Thurstone's (21) original college students, it failed to emerge consistently when similar tests were applied to schoolchildren, to whom such tests offered greater difficulty. It seems possible, again, that the word fluency factors, *W* or *f*, are differentiated from the verbal factor, *V*, largely by the ease of the verbal items involved, since several factorial investigations have failed to establish a clear separation between them. This is bound up with the whole question of the existence and nature of a speed factor in intelligence and other mental tests. For if such a factor is prominent it will presumably affect the performance of less able testees on an ordinary time-limit test more than it affects the more able ; and it may affect the later items of such a test more than the earlier ones. In spite of the rejection of such a factor by Spearman and others, the investigations of Slater (16), Davidson and Carroll (4) and Tate (20) have proved beyond doubt that power and speed of work can be differentiated under appropriate conditions.

It might be thought that internal analyses which reveal heterogeneity among test-items can be disregarded, provided that scores on the test as a whole give useful predictions of some external criterion of educational or vocational success. But it

is precisely because of the evidence accumulated during the war of irregular predictions at different test-score levels that the present writer has become concerned about heterogeneity.

One study of 600 naval radar operators yielded product-moment correlations of ·26 and ·35 for Progressive Matrices and Arithmetic tests with training marks. But when the 4 per cent. of failures and 96 per cent. of passes were contrasted, the tetrachoric coefficients were +·30 and −·13 respectively, showing that regressions were non-linear. The abilities measured by the Arithmetic test, or by the criterion marks, must have altered at the bottom end of the range. Similarly the Morse Aptitude Test, based on discrimination of Morse patterns, gave moderate correlations with Morse training results among telegraphists when product-moment correlations were used, but was extremely poor at picking out failures. In contrast the Morse Learning test (described by Vernon, 24) worked better at the lower than at the upper ranges. There was seldom time to investigate the significance of non-linearity, but it was definitely proven in one comparison, by Dr. Fraser Roberts, of the Matrices test with the naval T2 battery.

Clearly there is room for much more detailed analysis of what our tests measure, for determining the extent to which all items depend on the same factor or combination of factors, and whether heterogeneity of content or form is responsible for variations in predictive value.

The Progressive Matrices test[1] was originally devised as a measure of *g*, but it usually obtains small spatial or *k*-loadings, among adults, as shown by the writer elsewhere (24, 26). Its five sets of 12 items employ similar material, but differ slightly in their approach. Raven (13) describes them as :

(*a*) Continuous patterns,
(*b*) Analogies between pairs of figures,
(*c*) Progressive alterations of figures,
(*d*) Permutations of figures,
(*e*) Resolution of figures into constituent parts.

Each set starts with very easy items, so that separate instructions are not required, and each finishes with moderately, or very, difficult items. Only one study—to the writer's knowledge—has been made of the separate sets. Rimoldi (14) factorized these and 14 other tests, which had been given to 138 children of mixed sex, and age-range 8 to 15 years. Seven factors were extracted and rotated, and as might be expected the results are merely confusing. However, the unrotated factors indicate common variance in the Matrices sets of some 55 per cent., together with bipolars, which differ in the sets, of some 10 per cent.

Spearman (19) and others have suggested that different subjects might answer the items by different methods, for example, synthetically or analytically. It seems likely also that some would resort to spatial imagery, and others might employ logical verbal analysis. Heterogeneity of method in the persons would, on Burt's reciprocity principle, presumably imply heterogeneity of item-content in the tests. It was decided in the present study both to analyse the items (not the sets) among themselves, and to compare each item with other tests which would provide measures of well-established factors—*g*, *v*, *n*, *k*, mechanical information and perceptual speed.

II. DATA AND METHODS

The following battery of tests was assembled by Science 4, the Air Ministry, and given experimentally in the summer of 1948 to 640 National Service male recruits, all aged close to 18 years. Grateful acknowledgements are due for permission to use the results. Fuller descriptions of the tests may be found in Vernon and Parry (27).

[1] The original 1938 version was used in this investigation. It has, however, since been claimed that the later version, 1947, Set II, already meets many of the weaknesses described below. The items are arranged in groups of four, instead of twelve, according (*a*) to argument and (*b*) to order of difficulty. It would be preferable for this test to be used in future large-scale investigations of adult intelligence.

Arithmetic A. Short problems.
Calcs. A and Calcs. C. Four rules and mechanical arithmetic.
V-4. Stephenson's mixed vocabulary, English, and reading comprehension test.
V-V. A new vocabulary and comprehension test.
Gen A-1. Crossing out wrongly spelled words.
Gen A-2. Comprehension of paragraphs.
SP 14. A.T.S. Spelling ; identifying the correct spellings from several misspelled versions.
SP 21. Army Instructions or Clerical test.
SP 119. Giving the readings of pointers on scales and dials, and the co-ordinates of points on rectangular and polar graphs.
Ins-A. Reading aeroplane panel dials.
Obs-C. Matching aircraft silhouettes.
Progressive Matrices. Raven's 1938 test, with 20-minute time limit.
G-5. Stephenson's Matrices test, modified, with multiple-choice responses.
K-6. Stephenson's spatial test, including paper formboard, copying figures onto dotted space, and drawing mirror images.
SP 4. National Institute of Industrial Psychology's spatial judgment test, Squares.
Group Test 80. N.I.I.P. test of recognition of shapes turned round or over.
SP 122. Army revision of Bennett Mechanical Comprehension.
Mec-B. Diagrams of machines, comprehension questions.
Mec-C. Multiple-choice mechanical information questions.
SP 117 E and M. Questions on everyday electrical and mechanical matters, open-response.

Table I gives the 20-minute Matrix norms believed to be correct for 18-year males in the general population, and the distribution in our sample.

TABLE I. PROGRESSIVE MATRICES DISTRIBUTIONS

Score	18-yr. Population Per cent.	R.A.F. Sample Per cent.
50+	10	17·8
46+	20	25·7
42+	20	25·6
36+	20	21·6
28+	20	8·7
27 and under	10	·6

Unfortunately the sample is considerably above average ; and, though it is as heterogeneous as those normally encountered in educational or vocational investigations outside the Army, the Standard Deviation of scores is only 6·18. In more representative Army groups the Standard Deviation is about 8·50. Not only does this reduce the common-factor variance, but it also means that few recruits fail the easier items. Actually in 30 items out of 60 the number failing amounted to less than 10 per cent.

Tetrachoric correlations can still be obtained between items with very low (or very high) failure rates, but they are extremely unreliable. Thus the S.E. of zero r, according to Kelley's formula (11, No. 211) is ·062 when both items are passed by about 50 per cent. of a group of 640. But for the correlation between items $C\,1$ (p = ·970) and $E\,12$ (p = ·041), it rises to ·223. This difficulty will always arise in a test covering a wide range of ability. It might be met by taking an abnormally heterogeneous population, or, following McNemar, by making several overlapping analyses among low-grade and high-grade populations. However, most tests are not, like Matrices, intended to cover Mental Ages from 8 up to superior adult, and their items do normally lie between 10 per cent. and 90 per cent. difficulty.

The 18 easiest items were omitted altogether ; and the p values (difficulties) of the remaining 42 appear in Table II, Col. 2. The S.E. of zero r between each item and an item whose p = ·50 is given in Col. 3. For any other pair of items, the S.E. of zero r is given by the product of these two entries, multiplied by 16·1. (For example, $C\,1$ with $E\,12$: ·123 × ·112 × 16·1 = ·223.) The relative difficulties parallel very closely those quoted by Slater (17) for a group of 571 Army recruits, the rank order correlation between them being ·946.

Scores on the 22 tests were split at the median. It might be preferable to correlate each Matrices item with tests split at the corresponding difficulty level, but this would have enormously increased the labour, since the tests themselves would then have had to be inter-correlated and factorized at every level, or at least at several different difficulty levels. It appeared better to maintain a single level for the tests in order to provide a stable frame of reference factors with which not only the medium, but also the very easy and very difficult, items could be compared.

Tabulation of the 2,016 correlations between 42 items and 22 tests was done by hand, since punching was not readily available. Tetrachorics were read off from graphs. In view of the large S.E.s, and in order to save time in what was intended solely as an exploratory investigation, most of the subsequent calculations were done to two decimal places only.

III. ANALYSIS OF ITEMS

It is usually unwise to factorize two ' universes ' of variables, such as test items and test scores, in a single matrix. Separate analyses were therefore made of the inter-item correlations, the inter-test and the item-test correlations. The first was carried out by the Centroid method, and the four factors extracted are shown in Cols. 4–8 of Table II. Communalities were guessed after totalling the columns at each stage ; neither successive approximation nor the highest r in each column was used.

It is difficult to evaluate significance, but if we may assume (with McNemar) that a factor is equivalent to a variable split at the median, then the figures in the third column of Table II give us the S.E. of zero r for first-factor loadings. For the sth factor, the S.E., according to Burt and Banks (2) is $\sqrt{\frac{n}{n-s+1}}$ times larger. The writer's practice is to regard a factor as significant if half the loadings exceed $2 \times$S.E. Here 19 out of 42 fourth-factor loadings reach this level. On Guilford and Lacey's (9) criterion also, the fourth factor is close to the borderline of significance. Hence it did not appear justifiable to extract a fifth or further factors. Nevertheless the total communality obtained, 39·7 per cent., is a conservative estimate ; and it is likely that more efficient techniques of analysis, such as Maximum Likelihood, might raise it to perhaps 42 per cent.

This analysis is, by itself, not very informative. Most of the item inter-correlations can be accounted for by Factor I, with 26·0 per cent. variance, but the bipolars cover another 13·7 per cent. If rotation was attempted to factors with positive or near-zero loadings, the general factor would largely disappear, and independent factors corresponding quite closely to Raven's sets of items ($A+B$, C, D, and E) would be obtained. Most American factorists would almost certainly conclude that each set is measuring a different ability. However the heterogeneity is not really so serious, for two reasons, the first being the attenuation of Factor I by the homogeneity of the population. The total variance of 39·7 per cent. can be corrected by Kelley's formula (11, No. 178) to the value expected in a population with Standard Deviation 8·50. It then becomes 68·2 per cent., and as most of the increase would occur in Factor I, the variance of this factor might reach 54·5, that is, more than double the obtained value. Note that this figure is just about four times 13·7. Hence the Matrices test appears to satisfy the criterion of homogeneity suggested above (unless the bipolars also swelled in a representative group). Nevertheless it is instructive to realize how small the general factor may be, and how relatively great the heterogeneity, in a test applied to a moderately selected population.

Secondly, it is probable that factors with all-positive loadings would not give a true picture of item structure, and that at least one factor should be bipolar. Consideration of the answer sheets and item-correlations strongly suggests that many subjects spend so much time on the more difficult early items that they scarcely

TABLE II. RESULTS OF FACTOR ANALYSES OF MATRICES ITEMS

Column 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Item	p	S.E. at $p = \cdot 50$	Centroid Item Factors					Functional Factor Content							Formal Factor Content					
			I	II	III	IV	h^2	g	n	v	Technical	k	Perceptual	h^2	Bipolar	General	Group B	Group C	Group DE	h^2
A 11	·753	·067	·28	−·08	−·17	−·19	·15	·16			·14	·30		·14	+·13	·03	·12			·03
12	·920	·090	·30	−·06	+·20	−·03	·14	·34						·12	+·12	−·10			·24	·08
B 6	·947	·102	·33	−·03	−·34	−·35	·35	·23			·21	·32		·20	+·14	−·06	·23			·07
7	·872	·079	·37	−·24	−·23	−·14	·27	·23	·17			·14	·23	·16	+·29	·15	·20			·14
8	·833	·095	·30	−·34	+·28	−·14	·30	·22			·37			·18	+·16	·15	·03			·05
9	·706	·071	·28	−·25	−·12	−·13	·17	·22						·05	+·27	·11	·37			·21
10	·925	·092	·49	−·32	+·06	−·47	·57	·30		·31		·30	·25	·34	+·30	·14	·51			·36
11	·795	·070	·40	−·19	−·40	−·29	·44	·23	·18		·13			·10	+·36	·27	·41			·36
12	·601	·063	·48	−·15	−·14	−·18	·31	·38				·29		·23	+·23	·11	·25			·12
C 1	·970	·123	·48	+·10	−·30	+·10	·34	·25				·41	·22	·28	+·10	·23		·36		·19
4	·912	·088	·50	+·17	−·26	−·06	·35	·37				·27	·20	·25	−·02	·09		·48		·25
5	·941	·099	·57	+·03	−·23	−·19	·42	·44			·26	·34		·38	+·05	·12		·20		·05
6	·842	·075	·30	+·25	−·22	+·07	·21	·37	·17					·17	−·15	−·09		·42		·21
7	·937	·098	·60	+·06	−·08	−·11	·38	·40				·34		·28	+·09	·38		·15		·17
8	·689	·065	·44	+·11	−·16	+·13	·25	·25				·14		·08	+·05	·30		·36		·22
9	·742	·067	·38	−·05	−·17	+·16	·20	·33						·11	+·04	·19		·22		·09
10	·471	·062	·38	+·15	−·21	+·15	·23	·33				·24		·17	−·10	·05		·03		·02
11	·331	·064	·30	+·13	−·27	+·23	·23	·19	·12	·15				·07	−·01	·14		·36		·15
12	·061	·098	·53	+·22	−·15	+·18	·38	·67				·28		·53	−·19	·00				·04

D	2	·927	·092	·59	−·22	+·22	+·34	·56	·30		·11	·12	·14	·45	·34	+·17	·20	·66	·51
	3	·923	·092	·55	−·41	+·14	+·29	·58	·36		·13		·31	·32	·35	+·28	·00	·69	·56
	4	·902	·085	·51	−·43	−·06	+·26	·52	·38			·15			·17	+·34	·49	·16	·38
	5	·948	·104	·64	−·28	+·34	+·19	·64	·44	·17		·12			·24	+·20	·17	·88	·85
	6	·897	·084	·69	−·36	+·09	+·07	·62	·56			·20			·35	+·27	·51	·10	·34
	7	·800	·071	·57	−·36	+·16	+·01	·48	·34			·12			·13	+·26	·54	·22	·41
	8	·802	·071	·61	−·18	+·14	+·07	·43	·35		·13		·24		·20	+·15	·28	·52	·37
	9	·637	·063	·54	−·22	+·04	−·12	·36	·40			·14	·24		·20	+·13	·27	·14	·11
	10	·579	·062	·67	−·02	−·03	+·21	·49	·50			·22	·21		·34	+·07	·40	·08	·17
	11	·205	·070	·03	−·13	−·01	+·16	·04	·00					·15	·02	+·00	−·09	·30	·10
	12	·084	·089	·42	−·28	+·06	+·20	·30	·32		·27				·18	+·17	·34	·07	·15
E	1	·786	·070	·57	+·05	+·35	−·16	·48	·46			·14	·23		·28	−·20	·21	·15	·10
	2	·626	·063	·72	+·29	+·27	+·14	·70	·68			·13			·48	−·34	·40	·07	·27
	3	·643	·064	·64	+·18	+·08	−·18	·48	·61			·17	·12		·42	−·23	·42		·23
	4	·439	·062	·75	+·25	+·11	+·07	·64	·70			·14	·11		·52	−·26	·34	·06	·19
	5	·444	·062	·77	+·28	+·08	−·01	·68	·70						·49	−·21	·81		·70
	6	·369	·063	·62	+·31	+·08	+·05	·49	·59			·12	·15		·39	−·31	·40		·26
	7	·111	·082	·44	+·53	+·19	−·06	·52	·41		·11		·35		·30	−·45	—		·20
	8	·056	·101	·02	+·31	+·31	−·14	·21	·10		·27		·29	·38	·31	−·55	—		·30
	9	·282	·066	·48	+·35	−·10	−·18	·40	·45		·13		·36		·35	−·29	—		·08
	10	·242	·068	·70	+·32	+·04	−·09	·60	·70			·18	·19		·56	−·27	—		·07
	11	·204	·070	·63	+·25	+·14	−·16	·51	·63		·10		·30		·50	−·26	—		·07
	12	·041	·112	·27	+·28	+·35	+·03	·28	·60		·30				·45	−·52	—		·27
Variance per cent. ...				26·0	6·3	4·1	3·3	39·7	18·4	·3	1·0	1·4	4·3	1·6	27·0	5·9	7·5	9·2	22·6

Functional Content of Item-Factors	I	·86		·13	·10	·33		·88
	II	·16	·07	·11		·22	·11	·10
	III			·26	·08	−·15	·17	·13
	IV		−·15		·09	·28		·11
Variance per cent. ...		19·4	·1	·8	·3	3·5	·2	24·4

have time to tackle Sets *D* and *E*, whereas others guess freely and have more time to score on the later sets. If so, some groups of items really are negatively correlated with others. Doubtless this phenomenon arises largely because the test is applied with a time limit, which was not intended by its author. The moral is clear : tests given with insufficient time should have the items arranged in order of difficulty throughout, and if the items fall into distinctive sets, they should be arranged in omnibus or cyclic form. Otherwise the total scores of some subjects will be based to too great an extent on some of the sets, and of other subjects on other sets.

There is unfortunately no direct means of telling which bipolar originates in this manner. Order of difficulty gives some clue, although the subjective difficulty felt by the average testee does not necessarily correspond closely to the proportions of failures, $1-p$. By rotating Factors II and III through $\tan^{-1} \cdot 5$, a bipolar is obtained whose loadings give a tetrachoric correlation of ·73 with p. This is the best approximation we can get to a difficulty factor. It is shown in Col. 16 of Table II.

We cannot decide, either, whether our factors derive from the form or the functional content of the items. Probably all of them are mixtures. As already pointed out, the bipolars differentiate Raven's sets ; but it may well be that these sets differ in function (e.g., dependence on k) as well as in form. Although Factor IV yields mixed positive and negative loadings for at least two sets, *C* and *E*, we cannot assume that this represents differentiation by function. It may merely separate the more from the less difficult items within these sets. Thus the six most difficult items in Set *C* all have positive loadings and three of the four easier ones have negative. Similarly Items *E* 7–11 are all negative.

How far do general and bipolar variances differ with difficulty ? Table III gives the results for the easiest, medium, and most difficult items.

TABLE III. FACTOR VARIANCES OF EASY AND DIFFICULT ITEMS

Factor	I	II	III	IV	h^2
22 easy items, $p = \cdot 7$ to 1·0 ...	23·6	5·8	5·3	4·3	39·0
11 medium items, $p = \cdot 3$ to ·6 ...	35·2	4·3	2·5	2·2	44·2
9 difficult items, $p = 0$ to ·2 ...	20·5	9·8	3·5	2·1	35·9
All items	26·0	6·3	4·1	3·3	39·7

Apparently the medium ones have the greatest homogeneity and least bipolar variance. The easy and difficult ones both have higher than average bipolar content, but their low Factor I content is even more noticeable. The tendency for extreme items to be weaker than medium ones (even when consistency or validity is determined by tetrachoric or biserial correlation) was pointed out in the writer's survey of item-analysis techniques. The figure for the most difficult Matrices items is especially low because they include two items that are almost useless for adults, *D* 11 and *E* 8.

Any information about test items, based on general factor loadings, could of course be obtained much more simply by the stock item-analysis techniques. On the whole, then, factor analysis of items by themselves hardly seems worth while, provided that the internal consistency or reliability of the test is sufficiently high to ensure the dominance of a general factor. The bipolars have told us nothing, except that it is probably a mistake to give Matrices with a time limit. Conceivably they would be more useful in the case of a test where their meaning could be more readily identified by studying the content of the items they differentiate. Yet even in McNemar's Terman-Merrill analyses, interpretation was often difficult. An important reason for this is that the bipolar loadings are inevitably low in reliability when based on tetrachoric item-correlations.

IV. ANALYSIS OF TESTS

The inter-test correlations were analysed by group-factor methods, a previous centroid analysis of the same battery (slightly abbreviated) having revealed the general structure of the tests (Vernon, 26). *G* loadings were chosen to reduce to near-zero the correlations between sets of tests unlikely to have any group factor in common, viz. : Calcs. A and C ; G-5 and Matrices ; Mec-C, 117 E and M. Similarly V-4, Gen A–1, and SP 14 should be independent (apart from *g*) both of G-5 and Matrices, and of K-6, SP 4, and Group Test 80. The *g*-loadings of other tests, such as Obs-C or Scale Reading, were obtained by summing their correlations with the block of tests to which they were least related, and dividing by the summed *g*-loadings of this block. Admittedly there is an element of subjective judgment in this procedure, but it was guided both by the centroid analysis and by many previous analyses of most of the tests in comparable populations.

TABLE IV. FACTOR ANALYSIS OF TESTS

Test	*g*	*n*:ed.	*v*	Technical	*k*	Perceptual	h^2
Arithmetic A	·78	·44	(·16)	(·09)			·836
Calcs. A	·78	·59					·956
Calcs. C	·77	·53					·874
V-4 Verbal	·75	(·13)	·47				·801
V-V Vocabulary and Comprehension	·63		·58	(·13)			·750
Gen A-1 Spelling	·67	(·19)	·49				·725
Gen A-2 Paragraph Comprehension	·70		·38				·634
SP 14 Spelling	·58	(·32)	·69			(·15)	·937
SP 21 Instructions	·84		·28			·38	·928
SP 119 Scale and Graph Reading	·88	(·09)		(·19)		·24	·876
Ins-A Dial Reading	·90		(·13)			·26	·895
Obs-C Matching silhouettes	·53			(·19)	(·34)	·22	·481
Progressive Matrices	·84				(·27)		·779
G-5 R.A.F. Matrices	·90				(·28)		·888
K-6 R.A.F. Spatial	·74			(·13)	·64		·975
SP 4 Squares	·51			(·21)	·38		·448
Group Test 80	·56			(·23)	·53	·12	·662
SP 122 Practical Problems	·56		(·18)	·34	(·43)	(·28)	·725
Mec-B Mechanical Comprehension	·44		(·39)	·51	(·18)	(·27)	·711
Mec-C Mechanical Information	·62		(·28)	·50			·712
SP 117 E Electrical Information	·59		(·28)	·47			·647
SP 117 M Mechanical Information	·50		(·30)	·59			·688
Variance per cent.	48·9	4·5	9·0	6·4	6·0	2·3	77·1

After removing *g*, the prominent residuals between the three arithmetic tests gave the *n* (number) loadings shown in Table IV, by Spearman's general factor formula. Certain other tests had smaller positive residuals with these three, and their *n* saturations are shown in brackets. Clearly this is as much an elementary education as a pure arithmetic factor. Spelling obtains a considerable loading, but the verbal comprehension tests V-V and Gen A-2 are free from it. Five verbal tests similarly yielded the next, *v*, factor and this entered to some extent into the five

mechanical tests. Although this finding is unusual, it is quite plausible, since the mechanical tests obviously involve the understanding of printed material.

The next factor probably represents mechanical information and technical education, since Arithmetic, Scale Reading, Obs-C and the *k*-tests all show small loadings. The three spatial tests have a prominent additional group factor, which is present in mechanical comprehension (SP 122 and Mec-B) though not in information, also in G-5 Matrices and Obs-C. Finally there is slight evidence to support Thurstone's factor of perceptual speed in clerical and other tests which involve rapid matching or identification of verbal, numerical, or pictorial material. Actually its make-up is more reasonable in this than in the previous, centroid analysis, and though its reliability is low, it is probably significant. By Burt and Banks' formula, with $n=5$ and $s=3$ (since *g* and 1 or 2 group factors have been extracted from its component tests), the S.E. of a zero loading is ·081. Four of the five loadings exceed twice this figure.

Table IV shows that *g* covers 48·9 per cent. of test variance and the five group factors together 28·2 per cent., total 77·1 per cent. Corresponding figures for the centroid analysis were 43·8 per cent., 27·4 per cent., and 71·1 per cent., though comparisons are hardly legitimate both because the centroid analysis combined together certain pairs of tests here analysed singly, and because it was based on product-moment coefficients. It may be noted that neither analysis gave any indication of formal factors, although the tests do include open-response and various types of multiple-choice items, and although some are predominantly ' power ' and others are ' speed ' tests. Probably the functional content quite outweighed any grouping according to form.

V. FUNCTIONAL CONTENT OF ITEM FACTORS

The overlapping of items and tests could be approached in a number of ways. Here, each test was first treated as an additional Matrices item, and its correlations found with the four centroid item factors. These correlations were then treated as those of an additional test in the group factor analysis. The results appear at the foot of Cols. 9–15 in Table II.

As might be expected, the first item factor is largely composed of *g*. Its loading of ·86 corresponds nicely with the loading of ·84 for Matrices total score in Table IV. But it embodies a considerable amount of *k* also, and shows loadings with *v* and the technical factor which are probably insignificant. Only 12 per cent. of this factor cannot be identified functionally.

The bipolar factors are more complex. Their group-factor loadings are mostly non-significant, though they range up to ·28. Had the factors been appropriately rotated, it is conceivable that one might have been fairly highly *k*-saturated, another might have embodied all the *v* or *n* content, and so on. But the total functional content is so small that it seemed hazardous to attempt this. The total content (in so far as it can be identified by our reference battery) is $·260 \times ·88 + ·063 \times ·10 + ·041 \times 13 + ·033 \times ·11 = 24·4$ per cent. The difference between this and the communality of the item factors, 39·7, presumably represents formal factors. That is to say, the combined variance of the spurious factor introduced by the imposition of a time limit, and of common factors in all items or in groups of items which have no identifiable functional meaning, amounts to some 15 per cent. (It may be rather larger if, as suggested on p. 5, the 39·7 is an under-estimate.) A further attempt to discover, and assess the size of, the component elements of this 15 per cent. is described in Section VII.

VI. FUNCTIONAL CONTENT OF SEPARATE ITEMS

The tests were divided into five blocks corresponding to the five group factors (omitting G-5 and Matrices total), and the *g*-content of each item found from the block with which it gave the lowest correlations. Its group-factor content was then determined from its residuals. Table II, Cols. 9–15, lists only those group-factor loadings which are at least as large as their presumed S.E.s.

As would be expected, the *g*-loadings parallel closely the Item-factor I loadings, but are mostly rather smaller, since only 74 per cent. (i.e., $\cdot 86^2$) of Factor I is *g*. Loadings on *k* range from ·41 to zero, hence we can almost certainly conclude that some items do demand more spatial visualization than others. Sets *C* and *E* tend to be more spatial than *B* and *D* ; and this could have been deduced from the *k*-loadings of the item bipolars (thus Factor II, in which Sets *C* and *E* have positive loadings, correlates ·22 with *k*). It does not seem possible to tell, from inspection, which kinds of items are likely to be most spatial.

Several items, particularly in Set *E*, obtain *v*-loadings. Though unexpected, it is conceivable that such items demand a species of logical analysis in which the verbally minded have an advantage. Seven out of the 11 most difficult items in the test are among these *v*-loaded ones ; this corresponds to an r_t of ·74 between difficulty and *v*. The *n*-factor shows small loadings for five items only. This may be chance, but inspection suggests that they may involve counting or other mathematical thinking to a greater extent than the rest of the items. The perceptual loadings are interesting since they are mostly (though not invariably) confined to very easy items—thus confirming Thurstone and Shaefer's hypothesis. Four of the eight are among the eight easiest items, yielding an r_t of ·63. Obviously this association is statistically unreliable, but it is suggestive. Conceivably if we had been able to analyse more of the easy *A* and *B* items, we should have found this factor playing a more prominent part.

The most puzzling result is the number of items showing technical-factor loadings. This ought not to occur, since the test as a whole is free from the factor. Possibly it arises merely because mechanical comprehension tests are *k*-saturated, and it would have been avoided had the *k* group factor been extracted before instead of after the technical factor. By no stretch of the imagination can items *B* 8 and 6, C 5, *D* 10 and 6 be seen to involve technical knowledge any more than other adjacent items. Most of the remaining loadings are very small, as also are the technical loadings of the item bipolar factors.

When the variance of *g* and the group factors in all the items is totalled, it amounts to 27·0 per cent. That it should exceed 24·4 per cent. (the functional content of the item factors) is quite correct, since the specificity of an item might well consist, in part, of *g* or of some familiar group factor. This would explain, also, why some of the communalities in Col. 15 exceed those in Col. 8, though probably the main reason is, simply, unreliability.

Perhaps the chief finding of this section is that the functional content of Matrices items (as distinct from their formal factor loadings) does not alter markedly as one progresses. There will be a slight tendency for very low scores to be based on perceptual factor to a greater extent than average scores, and a rather more marked tendency for very high scores to be *v*-loaded. Also, of course, the average *g*-saturation tends to rise from beginning to end. But generally speaking scores at all levels covered by this investigation represent much the same mixture of *g* and *k*. From this point of view the Matrices test probably shows rather better homogeneity of content than does the Terman-Merrill scale. On the other hand Terman-Merrill is likely to be much more free from formal factors, for these undoubtedly do have different effects on different levels of Matrices scores.

VII. FINAL ANALYSIS OF FORMAL FACTORS

Since it seemed impossible to resolve the original item factors any further, a fresh attempt was made by first extracting the *g* and group factor content from the original item inter-correlations. The residuals, which were presumably now linked

by formal factors only, were then subjected to centroid analysis and three factors extracted. Their variance was 8·9, 4·2, and 4·7, total 17·8 per cent. This is a little higher than the 15 per cent. previously allowed, but naturally so because the analysis has gone much further. Instead of four item factors, we have removed *g*, five group factors, and three more bipolars.

These factors are not shown here, since all attempts to rotate them into logical structure proved fruitless.[1] They did however suggest that the structure consists of :

(i) the bipolar difficulty factor ;
(ii) a general factor in almost all items, representing the test's specificity ;
(iii) separate group factors in the *B*, *C*, and *D* sets, and possibly two group factors in *E* 1-6 and *E* 7–12.

Here we have seven factors. It is hardly to be wondered at that the original four item factors, in which not only these but also the six content factors were amalgamated, were difficult to interpret.

That No. (ii) exists, i.e., that there is some positive overlap among all items over and above their common *g*, *k*, etc., content, was proved by studying the residuals of each item with Matrices total (after removing *g* and *k*). The mean variance of these residuals was 4·5 per cent.

Since neither centroid nor group-factor techniques can readily cope with a general + bipolar + group-factor structure, it was eventually decided to use the rotated Item factors II + III (cf. p. 6) as an approximation to the bipolar difficulty factor, and to remove this first. The residuals were then analysed by Burt's group-factor technique. *E* 7–12 no longer showed any common variance, and *E* 1–6 overlapped closely with Set *D*, hence three, instead of five, group factors were sufficient. Item *A* 11 fitted in with the *B* set, but *A* 12 more closely resembled the *DE* set.

The results appear in Cols. 17–21 of Table II. A few of the loadings are abnormal, for example, *E* 5's general factor saturation of ·81 and *D* 5's group factor of ·88. But at the tail end of such complex manipulations of rather unreliable inter-item correlations this is only to be expected. The general factor loadings for the 36 items agree quite closely with the Matrices total score residuals ($r_t = ·76$). What is more important, the summed communalities for the six functional and five formal factors agree quite well with those of the four item factors ($r_t = ·82$), though they tend to be over-large.

The variance of the difficulty bipolar is 5·9, of the general factor 7·5, and the group factors 9·2, total 22·6. If the two abnormal loadings, just mentioned, are omitted, the general and group factor loadings drop to 5·9 and 7·4, the total to 19·2. This is still too large. We know that the general factor variance should be near to 4·5, and that the total should not much exceed 15. But considering that we have isolated 11 factors in place of four, the exaggeration is not unduly great. Our main conclusions are that the bipolar + general + group factors give a fair approximation to the structure of the formal factors, and that these three components are roughly equal in variance in the average item. There may be other formal factors linking smaller sets, or pairs, of items, but they are probably of negligible size.

Broadly speaking then an average Matrices item is determined by : *g* 19 per cent., *k* or technical 5 per cent., *v* or *n* 1½ per cent., perceptual 1½ per cent., positive or negative difficulty factor 5 per cent., Matrices factor (common to all items but to no other tests) 5 per cent., Set group factor 5 per cent. The total is 42 per cent., item specificity 58 per cent. Had the test been given to a more representative group of adults, the last figure would drop to 32 per cent. and, so far as we can guess, most of the increase would have occurred in the *g*-variance. But as we do not know what selective variables are operating in our population, this is uncertain.

[1] The first was positive for all items except *E* 7–12 and *C* 12. It was presumably a mixture of (i) and (ii). But it also gave very high loadings to the *D* set of items, i.e., it represented the group factor (iii) *D*. This explains why it could not readily be resolved by rotation.

VIII. CONCLUSIONS REGARDING FACTOR ANALYSIS

This exploratory investigation has been crude and not very accurate, and it has been hampered by inadequate tests of significance. (For example it is quite impossible to judge whether Matrices items differ significantly in perceptual factor loadings.) But this would seem to be inevitable when tetrachoric coefficients of poor reliability are analysed, and when the number of variables is far too large for successive approximation to communalities, or for the application of more efficient iterative techniques. In extracting the centroid factors, doubts frequently arose as to which variables to reflect, and there was certainly more than one possible solution which satisfied the criterion of positive column totals (disregarding the communalities). Wrigley (30) has recently discussed this problem and concluded that the factorist should aim to maximize the variance of each bipolar by reflecting tests with small loadings in such a direction as to increase the column totals of tests with large loadings, even if this does not produce all-positive totals.

In spite of these drawbacks the study has thrown much interesting light on test construction problems in general, and has permitted for the first time an assessment of the relative variance of functional and formal factors in a widely used group test. Whether the information obtained is of sufficient value to justify the work involved when constructing a new test is more dubious. There one often tries out a hundred items to pick the 50 best, and factorization would be out of the question. The analysis of items by themselves tells us very little which could not have been learnt more simply from stock item-analysis techniques. It appears to be capable of revealing only major defects of structure like the difficulty factor in Progressive Matrices and the prominent variations of functional content in the Terman-Merrill scale.

The inclusion of a battery of reference tests in the analysis constitutes a powerful tool, though again one that is very time-consuming. At least three tests must be given for each group factor that may be involved. Moreover, very large numbers of testees are essential before we can be certain, say, that an item designed to measure g is overloaded with k. If the present writer had been commissioned to revise the Matrices test, he would be extremely chary of eliminating or altering any of the items on the basis of the results given in Table II. Probably he would regard as the weakest items those that showed the largest proportions of functional group-factor and formal-factor variance, in other words the smallest g-variance. But this g-variance could have been obtained very easily by ordinary item analysis against a short battery of tests providing a reliable external criterion of g. So that here again factor analysis seems superfluous.

One way of making such research more manageable would be to combine items into a limited number of groups, each group being as homogeneous as possible in form, difficulty, and content. For example, the four easiest, four medium, and four most difficult items in each of Raven's sets might be treated as single variables. It would still be necessary to use tetrachorics. But the analysis of 15 item-groups would be a very different matter from analysing 60 items. It would be of great interest to do this with some of the popular group tests which contain several distinct sub-tests, or which follow the cycle-omnibus pattern, for example, Otis, Simplex, Cattell, Moray House, or Duplex.

IX. SUMMARY

1. The application of factorial analysis to a set of test items (or to the sub-tests of a battery) will help to show, first, whether the content of the test is homogeneous, or whether different items or sub-tests measure largely distinct abilities. It is suggested that the first or general factor running through all the items or sub-tests should obtain at least four times the variance of all the significant bipolar factors (before rotation), when analysed in a representative sample, if the test is to be regarded as adequately homogeneous.

2. Secondly, it may indicate whether unwanted 'formal' factors are present in some or all of the items. These include factors corresponding to the form in which the item is set, speed factors, or a factor contrasting difficult with easy items, etc. Such factors probably play an unduly large part in most multiple-choice tests. They are contrasted with 'functional' factors, which have useful psychological meaning, such as *g* and the recognized group factors, *v*, *n*, *k*, etc. It is possible that some of the familiar factors such as *W* and *F* (fluency) and *P* (perceptual speed) are primarily formal.

3. Such analyses are extremely time-consuming when the number of items is at all large, and the conclusion is reached that they contribute little which could not have been learnt from stock item-analysis techniques, unless it be certain grosser defects of the test's construction. Items must be inter-correlated by tetrachoric technique, and the resulting coefficients and factors tend to be very unreliable. It is suggested that analyses of groups of items, homogeneous in form, difficulty, and content, would be more economical. Several overlapping analyses at different difficulty levels may be essential for a test covering a wide range of ability.

4. Functional and formal factors cannot be distinguished unless the items are analysed along with an extensive battery of reference tests which will identify the item *g* and group-factor content. This too is seldom practicable, and is hardly necessary if ordinary item analysis is carried out against an external criterion.

5. Nevertheless heterogeneity and formal factors constitute a serious problem, as shown by the non-linear regressions frequently noted in following up Army and Navy tests during the war. An exploratory study was therefore made of one widely used test, the Progressive Matrices (1938), given with 20-minute time limit.[1]

6. Forty-two of the more difficult items of this test and 21 other reference tests were analysed among 640 R.A.F. recruits—a group somewhat superior to the general 18-year male population. The test's homogeneity, as shown by the variances of general and bipolar item factors, is low in this group, but would be just adequate in a representative population. Items of medium difficulty tend to be superior to easy and to very difficult items.

7. Owing to the imposition of a time limit there is a prominent 'difficulty' factor contrasting the earlier and later items. Other formal factors appear to include a general factor in all items (corresponding to the test's specificity in an ordinary analysis of tests), and separate group factors in most of the main sets of which the test is composed. The variance of each of these three kinds of factor in the average item is estimated at close to 5 per cent.

8. The functional content of the items is analysed into *g* 18·4 per cent. and group factors 8·6 per cent., but the figure for *g* might be more than doubled in a representative group. Different items depend to different extents on *v*, *n*, *k*, technical, and perceptual speed group factors; but *k* is much the most important. *V* appears

[1] This research was completed before I knew that Miss Keir had carried out a somewhat similar investigation among Secondary Modern Schoolchildren (G. Keir, 'The Progressive Matrices as Applied to School Children,' this *Journal*, 1949, II, 140–150). The results of the two studies generally confirm one another. Thus the order of difficulty of items among adult recruits gives a rank correlation of ·94 with that among Keir's boys. Nevertheless some of the misplacements of items to which she draws attention are reduced among adults; for example the mean difficulty of *D* items becomes greater than that of *C* items. Although the general factor-saturations of particular items differ considerably at the child and adult levels, my second and fourth bipolars sub-divide the items into the same groups as Keir's second and third. In both researches the reliability, general factor-saturations and correlations with other tests are reduced by the restricted range of ability of the testees. But my correlations with eleven verbal spelling and arithmetic tests are somewhat higher, averaging ·63 (range ·54 to ·73), as compared with her average of ·44. And the internal consistency of the test, as shown by the Kuder-Richardson Formula 20, is ·85 as compared with the odd-even coefficient of ·76 among children.

to enter chiefly into the most difficult, and perceptual factor into the easiest, items. Early items are also less *g*-saturated than later ones. But on the whole the test does measure much the same combination of functions at all levels (within the range covered by this group of subjects).

9. If, in a test given with a time limit, the formal or functional content does vary significantly, it is essential that the items be thoroughly mixed (e.g., in cycle-omnibus pattern) and presented in approximate order of difficulty, otherwise different levels of scores will measure different combinations of abilities.

REFERENCES

1. Burt, C. (1939). 'The Latest Revision of the Binet Intelligence Tests.' *Eugen. Rev.*, XXX, 255–260.
2. Burt, C., and Banks, C. (1947). 'A Factor Analysis of Body Measurements for British Adult Males.' *Ann. Eugen.*, XIII, 238–256.
3. Burt, C., and John, E. (1942). 'A Factorial Analysis of Terman Binet Tests.' *Brit. J. Educ. Psychol.*, XII, 117–127, 156–161.
4. Davidson, W. M., and Carroll, J. B. (1945). 'Speed and Level Components in Time-Limit Scores: A Factor Analysis.' *Educ. Psychol. Measmt.*, V, 411–428.
5. Eysenck, H. J., and Crown, S. (1949). 'An Experimental Study in Opinion-Attitude Methodology.' *Int. J. Opinion & Attitude Res.*, III, 47–86.
6. Ferguson, G. A. (1941). 'The Factorial Interpretation of Test Difficulty.' *Psychometrika*, VI, 323–329.
7. Guilford, J. P. (1940). 'Human Abilities.' *Psychol. Rev.*, XLVII, 367–394.
8. Guilford, J. P. (1941). 'The Difficulty of a Test and its Factor Composition.' *Psychometrika*, VI, 67–77.
9. Guilford, J. P., and Lacey, J. I. (1947). *Printed Classification Tests.* Army Air Forces Aviat. Psychol. Prog. Res. Rep. No. 5. Washington, D.C.: U.S. Government Printing Office. Pp. 919.
10. Hertzman, M. (1936). 'The Effects of the Relative Difficulty of Mental Tests on Patterns of Mental Organization.' *Arch. Psychol.*, No. 197. Pp. 69.
11. Kelley, T. L. (1923). *Statistical Method.* New York: Macmillan. Pp. 390.
12. McNemar, Q. (1942). *The Revision of the Stanford-Binet Scale.* Boston, Mass.: Houghton Mifflin. Pp. 189.
13. Raven, J. C. (1939). 'The R.E.C.I. Series of Perceptual Tests: An Experimental Survey.' *Brit. J. Med. Psychol.*, XVIII, 16–34.
14. Rimoldi, H. J. A. (1948). 'Study of Some Factors Related to Intelligence.' *Psychometrika*, XIII, 27–46.
15. Shaefer, W. C. (1940). 'The Relation of Test Difficulty and Factorial Composition Determined from Individual and Group Forms of Primary Mental Abilities Tests.' *Psychometrika*, V, 316–317.
16. Slater, P. (1938). 'Speed of Work in Intelligence Tests.' *Brit. J. Psych.*, XXIX, 55–68.
17. Slater, P. (1948). 'Comment on "The Comparative Assessment of Intellectual Ability".' *Brit. J. Psych.*, XXXIX, 20–21.
18. Smith, G. M. (1933). 'Group Factors in Mental Tests Similar in Material or in Structure.' *Arch. Psychol.*, No. 156. Pp. 56.
19. Spearman, C. E. (1946). 'Theory of General Factor.' *Brit. J. Psych.*, XXXVI, 117–131.
20. Tate, M. W. (1948). 'Individual Differences in Speed of Response in Mental Test Materials of Varying Degrees of Difficulty.' *Educ. Psychol. Measmt.*, VIII, 353–374.
21. Thurstone, L. L. (1938). 'Primary Mental Abilities.' *Psychometr. Monogr.*, No. 1. Pp. 121.
22. Thurstone, L. L. (1938). 'The Perceptual Factor.' *Psychometrika*, III, 1–17.
23. Vernon, P. E. (1938). 'Intelligence Test Sophistication.' *Brit. J. Educ. Psychol.*, VIII, 237–244.
24. Vernon, P. E. (1947). 'Psychological Tests in the Royal Navy, Army and A.T.S.' *Occup. Psychol.*, XXI, 53–74.
25. Vernon, P. E. (1948). 'Indices of Item Consistency and Validity.' *Brit. J. Psych. Stat. Sec.*, I, 152–166.
26. Vernon, P. E. (1949). 'The Structure of Practical Abilities.' *Occup. Psychol.*, XXIII, 81–96.
27. Vernon, P. E., and Parry, J. B. (1949). *Personnel Selection in the British Forces.* London: University of London Press. Pp. 324.
28. Wherry, R. J., and Gaylord, R. H. (1944). 'Factor Patterns of Test Items and Tests as a Function of the Correlation Coefficient: Content, Difficulty, and Constant Error Factors.' *Psychometrika*, IX, 237–244.
29. Wright, R. E. (1939). 'A Factor Analysis of the Original Stanford-Binet Scale.' *Psychometrika*, IV, 209–220.
30. Wrigley, C. F. (1950). *A Psychological Study of Methods of Predicting Success in Flying Training.* Ph.D. Thesis, University of London.

A METHOD FOR COMPUTING PRINCIPAL AXES

By L. F. RICHARDSON

I. *General Explanation.* II. *An Example Begun.*

I. GENERAL EXPLANATION

The following paper describes a working method for calculating the principal components of a correlation or covariance matrix. For reasons which will presently be obvious the procedure adopted has been called ' purification.' A full account of the principles involved is published in the Royal Society's *Transactions* (Richardson, 1950). The present note is intended as a brief guide to the special case for which the given $n \times n$ matrix K is symmetrical. Let the latent roots λ of K be arranged so that $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \ldots \geqslant \lambda_n$; and let them correspond respectively to latent columns $P_1, P_2, \ldots, P_n$ by way of the definition

$$KP_j = \lambda_j P_j \qquad (j = 1, 2, \ldots, n) .$$

The method of purification is a generalization of a routine which is familiar to psychologists, and which they usually trace back to Hotelling's paper of 1933 ; that is to say, the computer begins with a guess $X\circ$ at a latent column, and he proceeds by iteration. In Hotelling's method $X\circ$ is multiplied repeatedly by K ; and thus P_1 is obtained. My generalization consists in multiplication by $K - \rho I$, where ρ is a guess at an unwanted latent root and I is the $n \times n$ matrix having units in its principal diagonal, and zeros elsewhere. The presence of ρ gives the computer freedom to manœuvre. For example, if he desires to remove P_j, and if he can put $\rho = \lambda_j$, then he can compute $(K - \lambda_j I)X\circ$, and this does not involve P_j. Chemical language gives suitable hints. Thus $X\circ$ can be regarded as a raw material containing a desired constituent, which might on occasion be P_n, together with a lot of miscellaneous dirt consisting of $P_1, P_2, \ldots, P_{n-1}$. The multiplier $(K - \lambda_1 I)(K - \lambda_2 I) \ldots (K - \lambda_{n-1} I)$ would remove all that dirt.

To test whether a column Z is a pure latent column, let each element of KZ be divided by the corresponding element of Z, and denote the n quotients by $KZ \div Z$. If the n elements of $KZ \div Z$ are equal to one another, then they are equal to a latent root, and Z is its latent column. This ' division-test ' is already standard practice. The only novelty is the notation $KZ \div Z$. There is no need to consider the convergence of infinite sequences.

The iterations are shortened if the raw material $X\circ$ is rich in the desired constituent P_j. It is customary to start from $\{ 1, 1, 1, 1, \ldots, 1 \}$. But this, though rich in some constituents, is poor in others. It is preferable to choose as the raw material the appropriate latent column of some matrix which resembles the matrix in question.

Rayleigh's mean is a notable detergent. If a column X consists mainly of the latent column P_j, with only slight contamination by the $n - 1$ other P's, then Rayleigh's mean $\tilde{X}KX / \tilde{X}X$ is often an excellent approximation to λ_j. Here the symbol $\sim$ is used to denote transposition.

A precept of the purification method is that the wise computer locates all the latent roots roughly, before proceeding to find any latent root accurately. The iterations begin with the ' initial groping,' which is a game of skill, and follow on with the ' fine purification,' which is a routine. For any one latent root, the stage

of groping ends, and fine purification can begin, as soon as all the *other* latent roots have been roughly located.

Some well-known facts may here be usefully recalled. There is a quadratic form in the normalized test scores which represents an n-dimensional surface over which the frequency is constant. The matrix of that quadratic form is the reciprocal of the matrix of correlations (as was shown by K. Pearson, 1896). It follows that the semi-axes of the hyperellipsoid of constant frequency are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}, \ldots, \sqrt{\lambda_n}$. We may accordingly expect that the least latent root λ_n cannot be negative ; for, if it were, then the frequency would be finite in some directions at infinity. This is, however, only an expectation, not a definite proof, for the following assumptions have intervened : (i) that the persons tested were infinitely numerous ; (ii) that their distribution was Gaussian-normal ; (iii) that the correlations were not rounded off.

Much interest attaches to $\sqrt{(\lambda_n/\lambda_1)}$; because, if this quantity is practically zero, then the hyperellipsoid goes flat ; so that $n - 1$ suitably chosen co-ordinates suffice to describe the distribution.

II. AN EXAMPLE BEGUN

For illustration let us take Burt's (1915) study of the correlations between assessments for eight emotional traits. Let us call the matrix of observed correlations K.

TABLE I. CORRELATIONS BETWEEN EMOTIONAL TRAITS (BURT)

$$K = \begin{bmatrix} 1{\cdot}00 & 0{\cdot}83 & 0{\cdot}81 & 0{\cdot}80 & 0{\cdot}71 & 0{\cdot}54 & 0{\cdot}53 & 0{\cdot}24 \\ 0{\cdot}83 & 1{\cdot}00 & 0{\cdot}87 & 0{\cdot}62 & 0{\cdot}59 & 0{\cdot}58 & 0{\cdot}44 & 0{\cdot}45 \\ 0{\cdot}81 & 0{\cdot}87 & 1{\cdot}00 & 0{\cdot}63 & 0{\cdot}37 & 0{\cdot}30 & 0{\cdot}12 & 0{\cdot}33 \\ 0{\cdot}80 & 0{\cdot}62 & 0{\cdot}63 & 1{\cdot}00 & 0{\cdot}49 & 0{\cdot}30 & 0{\cdot}28 & 0{\cdot}29 \\ 0{\cdot}71 & 0{\cdot}59 & 0{\cdot}37 & 0{\cdot}49 & 1{\cdot}00 & 0{\cdot}34 & 0{\cdot}55 & 0{\cdot}19 \\ 0{\cdot}54 & 0{\cdot}58 & 0{\cdot}30 & 0{\cdot}30 & 0{\cdot}34 & 1{\cdot}00 & 0{\cdot}38 & 0{\cdot}21 \\ 0{\cdot}53 & 0{\cdot}44 & 0{\cdot}12 & 0{\cdot}28 & 0{\cdot}55 & 0{\cdot}38 & 1{\cdot}00 & 0{\cdot}10 \\ 0{\cdot}24 & 0{\cdot}45 & 0{\cdot}33 & 0{\cdot}29 & 0{\cdot}19 & 0{\cdot}21 & 0{\cdot}10 & 1{\cdot}00 \end{bmatrix}$$

This matrix, K, has also been studied by Holzinger and Harman (1948, p. 176). My table differs from theirs only by the fact that I have put units in the principal diagonal.

Raw material is needed for purification. The experienced reader may have in stock all the latent columns of some arithmetical matrix which closely resembles the given K. If so, he would do well to take them as his raw material. Here I shall have to be content with a model which has only a distant resemblance to K, but which is so simple that it can be analysed thoroughly by algebra. It is the $n \times n$ matrix A, called A for ' average,' and defined to have units on the principal diagonal, and to have the same number α in each other site.

A standard method for analysing A is to post-multiply it by the column T, called T for ' trial,' and defined thus :

$$T = \{ 1, \omega, \omega^2, \omega^3, \omega^4, \ldots, \omega^{n-1} \}.$$

The curly bracket is used when what is really a column is set horizontal. It is then found that each of the n elements of $AT \div T$ is equal to $1 + \alpha(\omega + \omega^2 + \omega^3 + \omega^4 + \ldots + \omega^{n-1})$ if only $\omega^n = 1$. In order to provide latent columns in sufficient variety, ω is at first taken to be a complex number, and at the end of the theory the real and imaginary parts of T are separated. By this method the n latent columns come out automatically as a mutually orthogonal set. There is a single isolated root, $\lambda_1 = 1 + (n - 1)\alpha$, for which $T = \{ 1, 1, 1, 1, 1, \ldots, 1 \}$. The other $n - 1$ latent roots are each equal to $1 - \alpha$. For the special case $n = 8$ the latent columns are shown in Table II. They do not involve α, and so may be applicable to other 8-rowed matrices of correlations.

TABLE II. ALL THE LATENT COLUMNS OF THE AVERAGED MATRIX A

$T_0 = \{$	1,	1,	1,	1,	1,	1,	1,	1 $\}$
$T_1 = \{$	$\sqrt{2}$,	1,	0,	-1,	$-\sqrt{2}$,	-1,	0,	1 $\}$
$T_1' = \{$	0,	1,	$\sqrt{2}$,	1,	0,	-1,	$-\sqrt{2}$,	-1 $\}$
$T_2 = \{$	1,	0,	-1,	0,	1,	0,	-1,	0 $\}$
$T_2' = \{$	0,	1,	0,	-1,	0,	1,	0,	-1 $\}$
$T_3 = \{$	$\sqrt{2}$,	-1,	0,	1,	$-\sqrt{2}$,	1,	0,	-1 $\}$
$T_3' = \{$	0,	1,	$-\sqrt{2}$,	1,	0,	-1,	$\sqrt{2}$,	-1 $\}$
$T_4 = \{$	1,	-1,	1,	-1,	1,	-1,	1,	-1 $\}$

The above eight columns may be regarded as jagged waves having respectively the following numbers of crests in the run of eight elements : 0, 1, 1, 2, 2, 3, 3, 4. The suffix to T is the number of crests. Waves having the same number of crests differ only in phase ; and one of them is distinguished by an accent.

There is a preliminary expectation that the latent columns of the given matrix K may resemble those of the matrix A ; and that the resemblance may extend also to the latent roots, if in A we put α equal to the average of those elements of K which are not on the principal diagonal. Then $\alpha = 0{\cdot}4603$. For this special form of A we have one latent root equal to $1 + 7\alpha = 4{\cdot}22$, and seven roots each equal to $1 - \alpha = 0{\cdot}54$.

The raw material is now ready for testing. Each of the columns T belonging to A is to be pre-multiplied by K. The operation is easy because the elements of the T are simple. First,

$$KT_0 = \{\ 5{\cdot}46,\ 5{\cdot}38,\ 4{\cdot}43,\ 4{\cdot}41,\ 4{\cdot}24,\ 3{\cdot}65,\ 3{\cdot}40,\ 2{\cdot}81\ \}\ .$$

These numbers are also, by exception, equal to $KT_0 \div T_0$. They are not equal to one another ; so T_0 is not a latent column of K. Yet they are of the same order of magnitude. Rayleigh's mean $\widetilde{T}_0 K T_0 / \widetilde{T}_0 T_0$ is, by exception, equal to the simple mean, and comes to 4·22. It may be regarded as a first approximation to λ_1. Next

$$KT_1 = \{\ 0{\cdot}14,\ 0{\cdot}59,\ 0{\cdot}90,\ \ 0{\cdot}05,\ -0{\cdot}46,\ -0{\cdot}23,\ -0{\cdot}15,\ 1{\cdot}02\ \}$$

$$KT_1 \div T_1 = \{\ 0{\cdot}10,\ 0{\cdot}59,\ \infty,\ -0{\cdot}05,\ \ 0{\cdot}33,\ \ 0{\cdot}23,\ \ \infty,\ \ 1{\cdot}02\ \}$$

These ratios are alarmingly unequal. The infinities arise, however, only from zero elements in the divisor. In Rayleigh's mean the zero elements act, not as divisors, but as multipliers ; and so they do no harm. It is

$$\widetilde{T}_1 K T_1 \big/ \widetilde{T}_1 T_1 = 2{\cdot}64/8 = 0{\cdot}33\ .$$

The same method has been applied to the columns T_1', T_2, T_2', T_3, T_3', T_4. The estimates which Rayleigh's mean gives for the eight latent roots of K are

$\lambda =$	4·22	0·96	0·64	0·55	0·52	0·48	0·33	0·28
From	T_0	T_1'	T_2'	T_4	T_3'	T_3	T_1	T_2

The sum of the eight estimates is 7·98. The sum of the true λ is equal to the sum of the principal diagonal of K, namely, 8 exactly. The agreement is close, but must not be allowed to engender too much confidence in these preliminary estimates ; for none of the division-tests was anywhere near satisfied. The seven equal roots of the averaged matrix A appear to have spread out a bit now that A has been replaced by K ; but the mean of those seven roots remains unaltered at 0·54. Altogether the averaged matrix A seems to be justified as a useful beginning. The raw columns T can now be purified with the aid of the preliminary λ which they have provided.

To find the greatest root λ_1. The other roots appear to form a cluster, well separated from λ_1. This allows us to ' kill seven birds with one stone,' namely, to remove most of the $P_2, P_3, P_4, P_5, P_6, P_7, P_8$, which occur as impurities in T_0, by way of the single multiplier $K - 0{\cdot}54I$. Thus let P_1^2 denote a second approximation to P_1 defined by

$$P_1^2 = KT_0 - 0{\cdot}54\,T_0 \quad .$$

It is important to notice that the upper of the two affixes to P is not the index of a power, but is the serial number of an approximation.

It follows that

$$P_1^2 = \{\ 4{\cdot}92,\ 4{\cdot}84,\ 3{\cdot}89,\ 3{\cdot}87,\ 3{\cdot}70,\ 3{\cdot}11,\ 2{\cdot}86,\ 2{\cdot}27\ \}$$
$$KP_1^2 = \{\ 21{\cdot}55,\ 20{\cdot}97,\ 17{\cdot}92,\ 17{\cdot}46,\ 16{\cdot}45,\ 13{\cdot}72,\ 12{\cdot}59,\ 9{\cdot}68\ \}$$
$$KP_1^2 \div P_1^2 = \{\ 4{\cdot}38,\ 4{\cdot}33,\ 4{\cdot}61,\ 4{\cdot}51,\ 4{\cdot}44,\ 4{\cdot}41,\ 4{\cdot}40,\ 4{\cdot}26\ \}$$

The elements of $KP_1^2 \div P_1^2$ are much more alike than were those of $KT^\circ \div T^\circ$. Considerable purification has evidently occurred. Rayleigh's mean is

$$\widetilde{P}_1^2 KP_1^2 \,/\, \widetilde{P}_1^2 P_1^2 = 4{\cdot}425,$$

and is probably correct to the first or second decimal. If further accuracy is desired as to λ_1 and P_1, it can rapidly be attained by repeated multiplication by $K - 0{\cdot}54I$. However, as this paper is about method rather than results, I leave the routine for refining λ_1 aside as obvious, and turn instead to :—

Approximations to the least latent root λ_8 *and its latent column* P_8. The column T_2, which yielded the least estimate of λ, has been chosen as the raw material. The choice of ρ, in the multiplier $K - \rho I$, is discussed in the full account of the method. For reliability everywhere ρ must exceed $\frac{1}{2}(\lambda_1 + \lambda_8)$. If nothing is known about the positions of the intermediate roots, $\lambda_2 \ldots \lambda_7$, then the best choice of ρ, for a single multiplication, is $\rho = \frac{1}{3}(2\lambda_1 + \lambda_8)$, which comes to about 3·0. Repeated multiplication of T_2 by $K - 3I$ would undoubtedly yield P_8 ; but the process would be insufferably slow, because of the cluster of roots close to λ_8. It is better here to take advantage of the special information that there is no latent root between 1 and 4·2. In this gap the purifier need not be reliable. Two multipliers can be used in conjunction, one to remove P_1, the other to remove $P_2, P_3, P_4, P_5, P_6, P_7$. The multiplier $(K - 4{\cdot}43I)$ will remove P_1 neatly, and is reliable everywhere. For the purpose of removing $P_2, P_3, P_4, P_5, P_6, P_7$, the constant ρ must exceed $\frac{1}{2}(\lambda_2 + \lambda_8)$. For a single multiplication by $K - \rho I$ it would be suitable to take $\rho = \frac{1}{3}(2\lambda_2 + \lambda_8)$, which comes to 0·73 on the uncertain information available at this stage. Actually I began with $\rho = 0{\cdot}7$ and afterwards decreased ρ to 0·6 when it became clear that λ_8 was less than expected. The multiplier $(K - 0{\cdot}7I)$ is not reliable everywhere : it both cleans and dirties. When the amplitude of P_8 is taken as the standard, then the multiplier $(K - 0{\cdot}7I)$ diminishes the amplitudes of $P_7, P_6, P_5, P_4, P_3, P_2$, but greatly magnifies the amplitude of P_1. To repress P_1 multiplication by $(K - 4{\cdot}43I)$ is used. The combined multiplier $(K - 4{\cdot}43I)(K - 0{\cdot}7I)$ is reliable everywhere, except in the gap where there are no latent roots. If the computer desires to multiply by a high power of a matrix, obtained by repeated squaring after the manner of Hotelling (Holzinger and Harman, 1948, p. 166), it would be suitable for him to work out a high power of $(K - 4{\cdot}43I)(K - 0{\cdot}7I)$. Here, however, in order to illustrate the process, I have used multipliers that are linear in K. It is important to notice that the impurity P_1 cannot be removed once for all at the outset. This is impossible partly because 4·43 is an inaccurate estimate of λ_1. But even if λ_1 were known precisely, nevertheless small amplitudes of P_1 would creep in whenever the digits were curtailed, and these traces of P_1 would be magnified in the subsequent multiplications by $(K - 0{\cdot}7I)$. When $(K - 0{\cdot}7I)$ and $(K - 4{\cdot}43I)$ are used alternately, the impurity P_1 alternately increases and diminishes, and may hide the steady decrease in the other impurities. In these circumstances the computer needs to make sure that there are no mistakes in his arithmetic ; he needs also to understand the theory of purification, and to have faith in it. Table III shows that the first multiplication by $(K - 0{\cdot}7I)$ did harm as well as good, for it increased the estimate of λ_8 obtained by Rayleigh's mean. A subsequent multiplication by $(K - 0{\cdot}6I)$ had a similar effect.

TABLE III. SUCCESSIVE APPROXIMATIONS TO THE LEAST LATENT ROOT OF K

Estimates of Latent Column P_8		Estimates of Latent Root λ_8		
How Obtained	Symbol	By Division-test Least	By Division-test Greatest	By Rayleigh's Mean
From matrix A	T_2	−0·06	∞	0·28
$KT_2 - 0{\cdot}7T_2$	P_8^1	−3·6	∞	1·13
$KP_8^1 - 4{\cdot}43P_8^1$..	P_8^2	−1·11	0·91	0·13
$KP_8^2 - 0{\cdot}7P_8^2$	P_8^3	−2·52	0·054	0·11
$KP_8^3 - 0{\cdot}7P_8^3$	P_8^4	−0·174	0·212	−0·005
$KP_8^4 - 0{\cdot}6P_8^4$	P_8^5	−1·273	1·978	0·031
$KP_8^5 - 4{\cdot}43\,P_8^5$	P_8^6	−0·296	0·029	−0·012
$KP_8^6 - 0{\cdot}6\,P_8^6$	P_8^7	−0·038	0·310	−0·013
$KP_8^7 - 4{\cdot}43\,P_8^7$	P_8^8	−0·036	0·274	−0·017

I have no doubt that λ_8 is negative, and that it can be excused as practically zero. The last approximation to P_8 is

$$P_8^3 = \{\ 92{\cdot}5669,\ 49{\cdot}2177,\ -78{\cdot}3811,\ -26{\cdot}0161,\ -28{\cdot}1690,\ -26{\cdot}9881,\ -29{\cdot}8714,\ 4{\cdot}2068\ \}$$

To find the second latent root it would be suitable to multiply T_1' repeatedly by $(K - 4{\cdot}43I)(K - 0{\cdot}3I)$.

But perhaps enough has been done to illustrate the method of purification.

I am obliged to Sir Cyril Burt for various guidance.

REFERENCES

1. Burt, C. (1915). *Report of the British Association for the Advancement of Science*, pp. 694–696.
2. Holzinger, K. J., and Harman, H. H. (1948). *Factor Analysis*. Chicago: University of Chicago Press.
3. Pearson, Karl (1896). *Roy. Soc. Lond. Phil. Trans.*, A, CLXXXVII, 301.
4. Richardson, L. F. (1950), *Roy. Soc. Lond. Phil. Trans.*, A, CCXLII, 439–491.

NOTE.—At Dr. Richardson's request I append the figures I myself obtained for the factor-saturations (latent vectors before normalization) and factor variances (latent roots) for the two largest factors. These results may be of assistance to any reader who desires to try his own hand at Dr. Richardson's instructive method of calculation. The figures given below were obtained to suit the requirements of Karl Pearson's "method of principal axes"; they were computed by weighted summation with unity in the leading diagonal (as in Table I above).

Trait	1	2	3	4	5	6	7	8	*Latent Root*
Factor I	·961	·933	·800	·780	·731	·608	·557	·427	4·439
Factor II	·047	−·145	−·452	−·186	·369	·244	·677	−·464	1·132

C. Burt

A STATISTICAL STUDY OF THE RORSCHACH TEST

By AMYA SEN

Department of Psychology, University College, London

I. *Problem.* II. *Subjects and Tests.* III. *The Validity of the Rorschach Results :* (a) *Cognitive Traits ;* (b) *Orectic Traits.* IV. *A Factor Analysis of Rorschach Categories.* V. *Summary and Conclusions.*

I. PROBLEM

The Validity of the Rorschach Tests. The object of the investigation described in the following paper was twofold : first, to compare the validity of the results obtained with the Rorschach inkblots (*a*) when the test is scored according to one of the recognized Rorschach methods, and (*b*) when it is scored in accordance with the principles suggested by Professor Burt (10) ; and secondly to determine objectively, by means of a factorial analysis, the most appropriate classification of the various characteristics revealed by the test, and to compare this objective classification with the classifications of personality types which the test is said to reveal when interpreted subjectively by an orthodox Rorschach technique.

The inkblot test as standardized by Rorschach and his followers has been described as " the most widely used tool in diagnostic personality testing " (7, p. 85). Yet it has been strongly criticized by those whose work lies specifically in the field of test validation. Burt, for example, states that, " as ordinarily scored, it yields validity coefficients that are far lower than assessments for the same traits based on an interview and lasting for about the same length of time " (10). Cattell declares that " despite the attraction which the Rorschach test has exercised, especially on psychiatrists, it remains a mixture of ill-defined intentions, analogous to a patent medicine, devoid of clear-cut theoretical bases " (21). Eysenck is equally trenchant : " To many Rorschach experts this test seems to fulfil the functions which in physics would fall to a spectroscope, cyclotron, thermionic valve, and closed chamber all in one : to more cautious psychologists, it appears as a test whose reliability is low, whose validity has never been established, and whose subjective nature does not attract the scientific worker " (22, p. 213). And Goodenough, after pointing out that the ' formulæ ' for interpreting the scores seem largely to have been " developed on the basis of chance relationships," concludes, a little more cautiously, that " the Rorschach method has now reached a point where a more critical attitude is desirable " (25, p. 436).

Empirical Validation. In general the critics of the test lay stress on two main defects. In the first place, despite its widespread use, hardly any researches have been undertaken to measure its reliability and validity by recognized statistical devices. There are plenty of investigations showing that it ' discriminates ' between various groups—particularly between the neurotic and the non-neurotic, and between different types of neurotics. But a test may yield unquestionable evidence for ' discrimination ' (i.e., for a fully significant difference between the group-averages), and yet possess little or no value as a diagnostic guide. As Rapaport observes, " the published comparisons of blind diagnoses with clinical histories and symptomatology do not validate the test : they merely demonstrate its potentialities " (20, p. 86 ; cf. also Goodenough, 25, p. 436).

One of the objects of the following paper will therefore be to provide quantitative indications as to the *amount* of validity that attaches both to the separate categories and to the test as a whole.[1]

Theoretical Construction. In the second place, as Rapaport points out, " the tie-up of the test with theoretical development in general psychology and the psychology of personality has also lagged." The traits or characteristics which the Rorschach test claims to assess consist in the main of an improvised list of psychological qualities which Rorschach himself suggested nearly thirty years ago ; and these for the most part assume the popular but highly fallacious principles so often stigmatized in criticisms of the faculty doctrine as the ' naming fallacy ' and the ' formal fallacy.' In consequence, as even the Rorschach expert himself at times admits, there is considerable danger that " the tester, especially the novice," may be led to a " dream-book type of interpretation " (20, p. 88). Goodenough makes a similar complaint. " There is no obvious resemblance between the nature of the subject's responses and the personality traits from which they are alleged to spring. Many of the ' signs ' reported have stemmed for the most part from ' hunches ' " (25, p. 435).

Nevertheless, ' the scientific worker ' is scarcely justified in dismissing a widely used test outright, without at least attempting to discover what degrees of validity its assessments actually possess, or suggesting constructive proposals for its modification and improvement. With the view, therefore, to making some systematic contribution towards these two main problems, a number of researches have been planned and carried out in the laboratory of University College. Of these, the earliest, so far as the Rorschach test itself is concerned,[2] was that undertaken by Miss Kerr, who applied the test to 265 normal, maladjusted, and defective children (11). A number of recent theses, published and unpublished, based on children or more frequently on adults, have dealt either with the Rorschach test as such, or with the Rorschach test as one of several methods of assessing personality (cf. 17).[3]

II. SUBJECTS AND TESTS

Subjects. One common line of criticism has been to suggest that the alleged correspondences between test scores and personal characteristics might be of social rather than of psychological origin. It seemed therefore desirable to obtain results from persons brought up under somewhat different cultural conditions from those who have generally been the subjects in most of the researches hitherto published.

[1] The few investigations that have attempted to assess *validity* (e.g., by correlating Rorschach scores with an independent criterion) have been concerned almost exclusively with the assessment of intelligence. The chief exception is that of Dr. P. E. Vernon (12, 13). In Table I of my thesis (27) I have compiled a comparative summary of the figures obtained by different workers, and have shown that they are extremely conflicting.

The *reliability* of the scoring and of the test-results was not a primary object of the present research, though a small sample of the scripts was scored independently by Ramzy. The results are given later in this paper. Miss Hillman has demonstrated that the reliability of the different blots, judged by two sets of scores based on half the blots, yields only a moderately high tetrachoric correlation (0·60 to 0·75 : unpublished thesis). In regard to the reliability of different examiners, Ramzy and Pickard have shown that, provided the responses are marked by two psychologists who have been in personal consultation over general principles, the agreement is reasonably close (24).

[2] In several earlier investigations an inkblot test had been used in London schools. In these cases, however, a composite battery of tests was generally employed, similar to that used by Burt and Moore, in which the test-material included not only inkblots and spot-patterns, but also ambiguous or indeterminate pictures, unfinished drawings and stories, as well as tests of intelligence. Of these researches the most suggestive was that of H. L. Hargreaves on ' The Faculty of Imagination ' (*Brit. J. Psych. Mon. Sup.*, X, 1927) who used tests of this nature with 131 children in L.C.C. schools. ' Inkblots ' and ' indeterminate pictures ' both proved to be valuable tests for two aspects of imagination termed ' fluency ' and ' originality.'

[3] I should like to acknowledge my indebtedness to all those who have assisted me in my own investigation, and particularly to Dr. Charlotte Banks, who has given special help both in the statistical analysis and in the writing of my paper.

Accordingly, for the present study, 100 Indian students were selected ; these comprised 60 men and 40 women, aged for the most part between 20 and 25, working at various subjects either in London or in Cambridge.

Tests. Each subject was given an individual Rorschach test, two intelligence tests (Group Test No. 33, a verbal group test prepared by Burt for the National Institute of Industrial Psychology, and the Progressive Matrices test, a non-verbal group test prepared by Raven), and Cattell's test[1] of fluency (based largely on Hargreaves' work, cited above). At a later stage they also took two questionnaires intended to elicit relevant traits of personality (for details see 27).

Scoring of the Rorschach Test. (*a*) *Rorschach Method.* Rapaport has suggested that " the lines along which the examiner should organize the scores " may be described as the " dimensions " of the tests, and that, in marking responses, these " dimensions " should be treated " as though they were a system of co-ordinates within which the individual case can be located." The term itself is intended to imply that all the response-characteristics grouped within one ' dimension ' are " pertinent to some major aspect of the subject's psychological make-up, and independent of other major aspects " (20, p. 109). However, the classification of response-characteristics for these purposes is evidently based, not on any factorial analysis of the test-results, but on impressionistic ideas of the various " psychological functions of the subject " which the test is supposed to reveal.

The " general rationale of the psychological processes underlying Rorschach responses," we are told, depends on the hypothesis that " the examiner may see in the subject's reactions to the inkblots perceptual organizing processes which are fundamentally continuous with his perceptions in everyday life." If, for example, the subject is observant when looking at inkblots, then, it is assumed, he will be observant in ordinary life ; if his responses are largely concerned with the dark shadows in the blots, then, it is supposed, he will tend to be unduly preoccupied with the dark shadows in his own existence ; and so on. The actual content or material of his responses is treated as of little consequence : " only in very sick patients, or cases with acute problems, does the content become directly revealing " (*ibid.*, p. 102). It is the ' form ' of the subject's mental processes that constitutes the main basis of the scoring. The general ways in which he reacts in seeking to give some intelligible interpretation to the visual stimuli before him are regarded as a sample of the formal characteristics that distinguish his mental reactions in actual life.

Among German writers on individual psychology this mode of interpretation has always been widely prevalent : it is admirably exemplified in their method of deducing personal characteristics from the corresponding characteristics seen in the person's handwriting. Among English-speaking writers, on the other hand, from Thorndike onwards,[2] the notion of " forms of mental activity, which work alike in every kind of concrete situation," has been highly suspect. It is true that recent factorial studies have at times confirmed the existence of a few formal factors, common to certain types of mental process : but, with rare exceptions, their contribution to the variance has proved to be small—far too small for diagnostic purposes.

For the detailed nature of the various ' categories ' commonly adopted by Rorschach workers, and of the scoring procedure generally employed, the reader may be referred to the standard works on the subject (9, 18, 19). Here it will be sufficient to note that the scores are usually classified under four main heads : (i) *Quantity*, i.e., number of responses and time taken ; (ii) *Area* or *Location*, e.g., how far the response covers the entire blot, or only parts, or possibly just minor details, and how successfully the response organizes the main constituents into a logically articulated whole ; (iii) *Determinants*, i.e., references to (*a*) colour, (*b*) form, (*c*) shading, (*d*) movement, and particularly (*e*) the relative balance of each ; (iv) *Content*, i.e., whether the things seen in the tests tend to be original or banal, and particularly what proportion can be assigned to certain stock headings—classes which are regarded as more or less symbolic, human beings, other animals, blood, fire, mountains, caves, sexual objects, and the like, and whether they show marked repetition or perseveration : it will be noted that, even in assessing ' contents,' it is still the broad formal characteristics that are noted, not their specific nature or their direct implications.

In the following enquiry, in order to obtain scores according to approved Rorschach principles, Beck's procedure (18) was used, with two minor changes. A slight modification

[1] Cattell, R. B. (1948). *A Guide to Mental Testing*, pp. 177f.

[2] Thorndike, E. L. (1923). *J. Educ. Psychol.* ' Individual Differences and their Causes.' III, esp. pp. 365f.

of his '*Z*-score' was introduced along lines suggested by the present writer,[1] and Klopfer's 'form level score' (19) was used in addition, with a view to determining which procedure was most effective. As a check on the reliability of the scoring, Mr. Ramzy was good enough to score 25 of the scripts independently: the contingency coefficients for 'location' and the various 'determinants' averaged 0·86 and 0·90 respectively.

(*b*) *Burt's Method.* The principles underlying Burt's suggestions may be summarized in his own words as follows.

"Inkblots have long been used as test-material, both in experimental and in individual psychology. Before James' criticisms [2] put the term out of fashion, they were regularly employed to exemplify what neo-Herbartian writers called 'apperception.' 'Apperception' may be defined as the mental process whereby the contents of a particular sensory presentation, at first relatively indeterminate and meaningless, become clearer and more determinate, and so acquire for the perceiver some kind of *meaning*, in virtue of a fusion or synthesis with contents already more or less systematically organized in his mind.[3] It is the fashion nowadays to describe the effects of this process by a popular metaphor from the magic-lantern: the perceiver is said to 'project' [4] his ideas and feelings in his mind on to the stimulus he is perceiving—an analogy which easily leads to extremely naïve, crude, and unjustifiable deductions. After all, thanks to the experimental work of Sharp,

[1] 27, p. 75. It seems more in accordance with the principles of scoring adopted by Beck and other Rorschach testers to regard the following types of response as not falling within the scope of Beck's *Z*: responses in which a small detail has suggested a comment bearing no genuine relation to the whole blot, responses consisting of combinations of small and large details where these are plainly inaccurate, and finally, vague, indefinite responses, such as 'maps' or the like. All these would receive a *Z*-value according to the official instructions for Beck scoring. In my own investigation both the unmodified and the modified *Z*-score were calculated and compared; but, as will be seen from the final results, the latter seemed unquestionably preferable.

[2] James, W. *Principles of Psychology*, II, pp. 107f. James quotes a long passage from Steinthal, who has perhaps given the most thorough analysis of the 'apperceptive processes' and their possibilities. Steinthal points out that, in ordinary everyday life, our commonest clues to a man's character and their place in life are furnished by his "mode of apperceiving." James comments that he himself finds Steinthal's analysis 'simply burdensome.' But an analytic study of the different processes involved is surely preferable to lumping them all together as comparable with the simple and direct 'projection' of a slide on to a screen.

[3] Stout, G. F. (1899). *Analytic Psychology*, II, ch. viii; (1913), *Manual of Psychology*, pp. 147f. As examples he cites inkblots, puzzle pictures, and clouds or constellations in the sky. Seashore's definition is similar to Stout's: "apperception designates the fusion of the presentative with the representative processes, including all their affective and conative processes, in the moment of a new experience. It is the grasping of a new experience so as to give it meaning and clearness. The interpretation is always personal" (*loc. cit. inf.*, p. 144). It may perhaps be added that, since the above was written, the term 'apperception' seems to be coming once more into favour: thus Bell commonly translates Rorschach's terms *Erfassung* and *Wahrnehmung* by 'apperception' (e.g., 23, pp. 88 and 190), and describes the 'determinants' as indicating 'quality of apperception.'

[4] Elsewhere Burt points out the widely different kinds of mental process that the ambiguous term 'projection' is made to cover. Bell, Rapaport, and Goodenough, for instance, all treat the Rorschach test as a typical "projective procedure"; but they explain the term in very different ways. Goodenough gives three explanatory examples: (i) "We say our soup is cold": here we attribute the "sensation in our mouths" to the soup outside us; (ii) "the office-manager, during an attack of indigestion, raises a tempest over a misfiled letter; he projects his feelings on to the persons and things surrounding himself." But he does not actually attribute his angry *feelings* to the letter or the clerk. (iii) "In some forms of mental disease projection is carried to the point of delusion: the patient sees demons or hears voices of angels." Here hallucinatory reality is attributed to a visual or auditory image—again a different process. Rapaport, on the other hand, attributes the projection not to the person but to the test: "the concept of projection in projective procedure is formed on the pattern of projector and screen." No wonder Bell observes that "no clear or common definition of what is meant by 'projective' has as yet appeared among those who use these methods." He is quite wrong, however, in stating that "projection was first used in a psychological sense by Freud." The idea of such a process has always been a favourite one with physiological and medical writers; and was criticized long ago by James (*Principles*, II, p. 31). As Ward has put it, "those who talk about 'projection' have usually committed the prior fallacy of 'introjection'."

Pillsbury, Bartlett, and others,[1] we do know a good deal about the varied processes involved in apperceiving inkblots, though writers on the Rorschach tests make little or no mention of it.

"Older books on experimental psychology commonly started with a laboratory experiment of this type to demonstrate to the student how much of what he thinks he sees is really an unconscious contribution of his own mind, and how characteristically the interpretations offered differ from one observer to another. 'To a hunter,' it is said, a particular inkblot 'suggests a beaver or a woodchuck; to a mason, it is a trowel; to a keeper of pets an Angora cat; to the naturalist a hedgehog or flounder, according as his mind has been most directed to land animals or to fish.'[2]

"It follows that, by comparing the particular ways in which different individuals *apperceive* the same series of 'equivocal stimuli,' we may often secure a detailed insight into their habitual vocabulary and cultural background, their dominant interests and ideas, their past history and experiences, and even their emotional attitudes and complexes. In all such cases, however, it is the *contents* of the 'apperceptive contributions,' as disclosed by the contents of their interpretations, that provide us with the requisite data, rather than the *formal* characteristics of their perceptual or associative processes. . . . In summarizing the inferences drawn according to this principle it will be natural to classify the traits in accordance with the usual type of psychographic scheme, elaborated by the workers in the field of individual psychology, instead of inventing some new scheme and some special list of categories, expressly for the purpose of this particular test."

In the present enquiry the assessments were limited to those particular traits which previous trials had shown could be most reliably rated. Accordingly, data extracted from responses were classified under the following heads: Intelligence, Verbal Ability, Imagination, Perception of Relations, General Emotionality, Extraversion-Introversion, Assertion-Submission, Cheerfulness-Depression, Sociability, Anxiety, Neurotic Tendencies, and Extent of Vocational and Cultural Interests. These largely correspond with mental factors already established in the fields of cognitive or orectic psychology. The numerical ratings were obtained by counting up the number of relevant items in accordance with the scheme suggested by Burt; the score for each trait was then converted into terms of a 15-point scale, running from E− to A+.[3]

Criteria. Each of the subjects tested were rated independently by two judges for the traits enumerated above. The judges were friends or close acquaintances of the subjects whom they rated, and had lived under the same roof with them for at least two years. It proved impossible to find a single pair of judges who were well acquainted with each one. Three pairs were able to rate groups of 15, 13, and 10 subjects respectively; the rest rated somewhat smaller groups. Each was asked to assess the subjects in his (or her) group for one trait at a time; and the assessments were made on a 15-point scale, similar to that used for rating the Rorschach scores.

Reliability of Judges. To estimate the reliability of the ratings, an overall correlation was first calculated between the two ratings thus secured for every trait, regardless of any differences in the judges. The coefficients ranged from 0·510 to 0·791, and averaged 0·712. Separate correla-

[1] Pillsbury, W. B. (1896). 'A Study of Apperception.' *Amer. J. Psych.*, VII, pp. 350–393. Sharp, Stella (1899). 'Individual Psychology: a Study in Method.' *Ibid.*, X, pp. 329–391. Bartlett, F. C. (1916). 'An Experimental Study of Some Problems of Perceiving and Imagining.' *Brit. J. Psych.*, VIII, pp. 222–266. Cf. also Hicks, Dawes (1914). 'The Nature and Development of Attention.' *Ibid.*, VI, pp. 7f.

[2] Quoted from Seashore, C. E. (1908), *Elementary Experiments in Psychology*, Ch. XII. Cf., also Witmer (1902), *Analytical Psychology: A Practical Manual.* Ch. I, esp. Experiments III and IV on 'The apperceptional contribution in perceptions of equivocal stimuli' and 'Variable and individual apperception.' In his own experiments on individual differences at Liverpool, Burt used, in addition to material taken from Witmer and Sanford, ambiguous pictures and artificial asymmetrical blots (uncoloured and coloured), which he found in later experiments slightly superior to the symmetrical natural blots of the kind used by Rorschach: both the average reliability (0·41) and the average validity (0·21) were low; but sex differences were easily demonstrated (*J. Exp. Ped.* (1911), I, p. 381; *Brit. J. Educ. Psych.* (1945), XV, pp. 116f.). To control the presentation, and investigate the processes commonly involved, he employed McDougall's 'portable tachistoscope' (which can be adapted to the ordinary lantern for group presentation). Subsequently, as Myers' assistant at Cambridge, he started a more intensive study of 'the general apperceptive processes'; and this later work formed the basis of the present scheme.

[3] The details of the method are described and illustrated in unpublished *Laboratory Notes on the Rorschach Test* (10). For the psychographic schedule employed, see 26 and refs.

tions were then calculated between the two ratings for the three largest groups, each rated by the same pair of judges. The figures obtained ranged from 0·450 to 0·855, and averaged 0·642. The general size and range of the reliability coefficients is thus not very different in the two cases. Since by mixing ratings secured from different judges the size of the coefficients is not significantly increased, we may assume that their standards were pretty much the same. In general, traits which are openly manifested in everyday intercourse, such as intelligence, sociability, intellectual and cultural interests, verbal ability, anxiety, and cheerfulness, were assessed with the highest reliability (about 0·75 to 0·80); traits like imagination or general emotionality were assessed with the lowest reliability (about 0·50 to 0·60).

III. THE VALIDITY OF THE RORSCHACH RESULTS

(*a*) *Cognitive Traits.* (i) *General Intelligence.* To secure a validating criterion for general intelligence, the scores in the two intelligence tests were reduced to standard measure, and averaged. These average scores were then correlated, first, with the judges' ratings for intelligence, secondly, with the ratings for intelligence obtained from the Rorschach responses with Burt's procedure, and thirdly, with certain 'categories,' scored by Beck's method, which are commonly claimed by orthodox Rorschach workers to be indicative of intelligence.

The correlation between the intelligence tests and the judges' ratings was, for the men and women together,[1] 0·671. The correlation between the tests and the ratings given by Burt's content method was 0·765 (men, 0·738; women, 0·792). The correlations between the tests and the several Rorschach categories, scored by Beck's method, are set out below in Table I. In each case the coefficient was calculated by the full product-moment formula.

TABLE I. GENERAL INTELLIGENCE: CORRELATIONS BETWEEN INTELLIGENCE TESTS (CRITERION) AND RORSCHACH CATEGORIES

CATEGORY	CORRELATION	
	Men	Women
Responses to Whole Blot (*W*%)	−·060	·062
Large Details (*D*%)	·102	·052
Large Unusual Details (*D* + *Dd*%)	·046	·056
Good Form (*F* + %)	·457	·436
Movement Responses (*M*%)	·408	·415
Beck's *Z*-score	·221	·403
Modified Beck's *Z*-score	·409	·415
Klopfer's Form Level Score	·500	·586
Total Number of Responses	·055	·053
Animal Responses and Animal Detail (*A* + *AD*%)	−·365	−·190
Popular Responses (*P*%)	·248	−·143

[1] Rorschach himself and most other Rorschach workers hold that there are no perceptible differences between the sexes, and therefore do not tabulate results from men and women separately. Burt and Moore, however, found marked sex-differences with their tests, particularly among children; and considered these to be analogous to the different sex interests reflected in other tests of apperception and association. Thus, in experiments carried out during 1914–18, "where the blots or the pictures are coloured, the girls were far more prompt to notice the fact than the boys; on the other hand, the boys made more references to action. Among the boys there were nearly three times as many references to the war as among the girls. . . . Among adults, however, except for occupational and domestic interests, sex-differences were much smaller." Most other workers with inkblots have also reported distinct sex-differences, at least among children of school age (cf. 4 and refs. 8, 11).

In my own groups, perhaps owing to their relatively small numbers, perhaps owing to their different cultural background, sex-differences in the Rorschach scores were in most instances comparatively slight. It therefore seems legitimate to pool the results from the two sexes in computing the correlations. Wherever there is any doubt on this point, I have tabulated coefficients for the two sexes separately.

Coefficients of 0·255 and 0·330 mark the 5 per cent. significance level for the men and the women respectively ; the significance level for the average for both sexes is 0·196. It will be seen that the results obtained from the two sexes separately tend to confirm each other. Four categories give correlations over 0·400 for both sexes, namely, the modified Beck score, Klopfer's form level score, percentage of good form, and percentage of movement. None of these, however, approaches the figures obtained by the content method.

(ii) *Other Cognitive Traits.* The judges' ratings for Imagination, Verbal Capacity, and Vocational and Cultural Interests were correlated with the ratings for the same traits obtained by Burt's method ; and the judges' ratings for Imagination were correlated with two Rorschach categories claimed to be indicative of this quality, namely, percentage of 'movement' and of 'form' responses. (The Rorschach scoring gives nothing comparable to the ratings for the other two traits.) The coefficients are shown below in Table II. Since here there are no appreciable differences between the figures for the men and women, the results for the two sexes have been pooled into a single group.

TABLE II. COGNITIVE TRAITS : CORRELATIONS BETWEEN JUDGES' RATINGS AND RORSCHACH RESULTS (*a*) WITH BURT'S METHOD, (*b*) WITH BECK'S METHOD

TRAITS	A. Content Method	B. Beck's Method	
		Percentage of Movement	*Percentage of Form*
Imagination	·445	·127	—·212
Verbal Ability	·535	—	—
Intellectual Interests ..	·486	—	—

As before, with a group of this size a coefficient of $\pm 0{\cdot}196$ marks the level of statistical significance. For Imagination the content method of rating yields a much higher correlation than the standard procedures. For verbal ability and intellectual interests the results confirm the expectation that an analysis of content will provide a fair indication of these qualities ; and since this is such an obvious component of the material obtained, it seems unfortunate to omit it (as the orthodox Rorschach procedure does) from the inferences to be drawn from the test.

TABLE III. ORECTIC TRAITS : CORRELATIONS BETWEEN JUDGES' RATINGS AND RORSCHACH RESULTS (*a*) WITH BURT'S METHOD, (*b*) WITH BECK'S METHOD

TRAIT	CATEGORY	A. *Content Score*	B. *Category Score*
General Emotionality ..	Percentage of Colour Responses (*C*%)	·600	·198
Extraversion-Introversion	Percentage of Movement minus Colour Responses ($M - C$%)	·540	·061
Assertion-Submission ..	.. ..	·433	—
Cheerfulness-Depression	Number of Form-Vista plus Flat Grey Responses ($FV - FY$)	·519	·082
Anxiety	Number of Form-Vista plus Flat Grey Responses ($FV - FY$)	·624	·138
Sociability	Number of Human Responses (*H*%)	·479	·230
Cultural Interests ..	.. ..	·663	—

(*b*) *Orectic Traits.* We now turn to the estimates for affective and conative traits. Once again the ratings of the two judges for each student were averaged ; and the averages were then correlated with the Rorschach ratings as obtained by the two alternative procedures. The results are shown in Table III.

The highest correlation with the Rorschach categories is that between the ' number of human responses ' and the ratings for sociability. But the low figure (·230) demonstrates that, even in this field, the Rorschach categories do not yield very useful results.

Matching Method. It is sometimes admitted by Rorschach workers that the individual Rorschach categories, taken by themselves, are of but little use for diagnostic purposes. Rather, they say, it is the total ' psychogram,' the *Gestalt*, as it were, of all the categories, that alone will give a successful indication of the real *Struktur* of the individual's personality. To test this claim, the Matching Method provides the most appropriate statistical technique. A verbal character sketch is compiled from each Rorschach protocol ; and independent judges are then asked to match the names of the subjects with the several sketches submitted.

In the few researches on the Rorschach test where this method has been used, fairly high coefficients of agreement have commonly been obtained (23, p. 135). But there is some evidence to suggest that this may be due in part to limitations in the method itself. Vernon has pointed out certain drawbacks in the technique that are sometimes overlooked (14). He notes that the results (i) vary greatly with the heterogeneity of the group to be matched ; (ii) depend on the experience and insight of the judges ; and (iii) demand a good deal of experience and skill on the part of the interpreter who interprets the Rorschach responses and writes the sketches. One further drawback arises from the fact that only very small groups of subjects can be conveniently matched ; and it has often been objected, on purely *a priori* grounds, that any small and heterogeneous group could probably be matched with fair success, whether the personality sketches were based on Rorschach responses, common observation, or indeed almost any relevant procedure. Thus, in my own cases subsequent enquiries from the judges often showed that they had been frequently guided, not by the sketch as a whole, but by some specific clue which had been inadvertently included by the writer and which was not really dependent on the test-results.

In the present research the total number of subjects was split into small groups, five to seven in each ; and personality sketches were written from the orthodox Rorschach protocols. The results gave high figures : by Burt's formula the average coefficient was ·849 ; by Vernon's the combined figure was ·856. As a rough comparison, personality sketches based on the scoring by content were later read to judges of four of the groups. Here the coefficient was 1·000, i.e., every one of the subjects was matched correctly, although the sketches were decidedly short, and dealt with the cognitive traits quite as much as the orectic.

These figures, however, are not very trustworthy, partly owing to the difficulties in the matching technique already noted, and partly on account of the conditions of the experiment. For example, when writing up the sketches the investigator had already noted the contents of the responses. It was therefore difficult to be sure that the sketches, nominally based on a synthesis of the Rorschach category scores, had not also been influenced by attention to the content. Probably a more reliable way of approaching this problem will be to combine the Rorschach category scores so as to obtain weighted sums for the various traits that we are seeking to measure ; and for this purpose we must have recourse to the more rigorous methods of factor analysis.

IV. A FACTOR ANALYSIS OF RORSCHACH CATEGORIES

Correlations between Categories. As Burt has pointed out, with the wide variety of Rorschach scores, two modes of factorial analysis are available—the first, based on the more familiar method of ' correlating traits,' the second on the comparatively novel method of ' correlating persons.' The latter, however, though highly suggestive, is not only less usual, but also more elaborate, and would require a paper to itself. Here, therefore, I shall confine myself chiefly to results obtained by the better known procedure.

Tetrachoric correlations have been calculated between the main Rorschach categories, thirty-six in all.[1] The two intelligence tests were also included in the series. If we accept the common assumption that the standard error of the tetrachoric coefficient is approximately 1·50 times the standard error of a product-moment correlation, then we may regard coefficients above 0·300 as statistically significant.

A number of the correlations had minus signs. As a result, on summing the columns to extract the first factor, it was found that eight of the categories produced negative totals, namely, Popular Responses (*P*%), Responses to Whole Blot (*W*%), Animal Responses (*A*%), Anatomy Responses (*AN*%), Responses to Form (*F*%), Beck's *Z*-score, the Modified *Z*-score, and, finally, the Average Time taken to make the first response. It would, of course, be quite impossible to start with a bipolar factor containing 29 positive and only 8 negative saturations. Accordingly, to secure positive saturations throughout the signs for these seven categories were reversed.

TABLE IV. FACTOR-SATURATIONS FOR RORSCHACH CATEGORIES AND INTELLIGENCE TESTS

Category	A. BIPOLAR FACTORS			B. GROUP FACTORS			
	I	II	III	*F*	*E*	*G*	*P*
Intelligence Test (Verbal)	·267	·516	—·152	·040	·590		
Intelligence Test (Non-Verbal)	·298	·675	—·085	—·011	·744		
Form Level Score (Klopfer)	·077	·805	—·189	—·260	·829		
Good Form (*F* + %)	·124	·731	—·170	—·240	·860		
Movement (*M*%)	·097	·597	·288	—·259	·800		
Few Anatomy (*AN*%)	·469	·510	·264	·235	·691		
Human Responses	·561	·362	·108	·378	·601		
Few Animal Responses (*A*%)	·595	·340	·143	·451	·546		
Architecture	·420	·112	·109	·417	·254		
Vocational	·475	·100	·103	·497	·219		
Form and Colour (*FC*)	·426	·088	·291	·384	·174		
Object	·619	·086	·050	·671	·245		
White Spaces	·546	·040	·008	·607	·159		
Few Form Responses (*F*%)	·163	·724	·288	—·248		·900	
Vista (three-dimensional)	·554	·364	·281	·400		·584	
Colour (*C*)	·142	·230	·623	·035		·469	
Surface Grey (*FY*)	·630	·172	·131	·525		·411	
Landscape	·748	·157	·288	·618		·521	
Abstract	·368	·113	—·415	·398		·336	
Clothing	·480	·088	—·004	·455		·178	
Few Popular Responses (*P*%)	·389	—·588	—·152	·495			·458
Human Details	·506	—·522	—·099	·513			·518
Proportion of Coloured to Uncoloured Responses	·166	—·306	—·109	—·077			·594
Low Average Response Time	·211	—·242	—·125	·327			·113
Clouds	·366	—·202	—·650	·446			·068
Animal Movement (*FM*)	·101	—·179	—·129	·007			·281
Colour and Colour Form (*C* + *CF*)	·344	—·125	—·560	·364			·077
Sex	·362	—·064	—·107	·342			·102
Cultural	·850	—·063	—·131	·780			·170
Low Modified *Z*-score	·135	—·998	·131	·128			·900
Low *Z*-score (Beck)	·429	—·702	·261	·372			·743
Bad Form (*F* —)	·377	—·650	·124	·545			·443
Few Responses to Whole Blot (*W*%)	·615	—·547	·510	·450			·694
Animal Details (*AD*)	·535	—·528	·036	·554			·514
Unusual Details (*Dd*)	·903	—·438	·075	·907			·473
Percentage of Large Details (*D*%)	·503	—·341	·344	·343			·505
Total Number of Responses	·962	—·266	·224	·897			·375
Plants	·482	—·046	·378	·400			·113

[1] The correlation table itself is too large to reproduce here. It will be found in the Appendix to my thesis (Table III, following p. 223).

1. *Bipolar Analysis.* The correlation table was first factorized by Burt's Method of Simple Summation, using successive approximation to secure appropriate communalities (16, pp. 461f.). Three factors, but no more, were found to be statistically significant, namely, one general factor, with positive saturations throughout, and two bipolar factors (cf. Table IV A). As has frequently been insisted (16), to be convincing any psychological interpretation of the statistical factors obtained in this way should, if possible, be supported both by internal evidence and by the evidence of external criteria. It will be convenient to examine the internal evidence first.

If, as is claimed, the Rorschach test covers the entire personality, and if further we accept the usual broad analysis of personality into two main aspects, cognitive and orectic respectively, then we might expect the two bipolar factors to represent *general intelligence* and *general emotionality* respectively, i.e., the one would contrast responses indicative of high intelligence with those indicative of low intelligence, and the other would contrast traits indicative of strong emotionality with those indicative of weak emotionality. As we shall see, owing apparently to the rather specialized diagnostic interests of those who have built up the method of scoring, this interpretation is not quite accurate. Further, the presence of a comprehensive *general factor* would seem to be somewhat unexpected to those who have hitherto offered a ' rationale ' of the test.

(i) The first factor obtained with this procedure is a ' general ' factor, having positive saturations ranging from 0·077 to 0·962. But, since so many traits have extremely low saturations, it contributes only 29 per cent. to the total variance.

From its very nature, a ' general ' factor denotes efficiency or favourable performance in all the traits in the series from which it is obtained. A positive measurement for this factor must therefore represent favourable performance in the Rorschach test, as scored by the orthodox Rorschach method.[1] Can we give this factor any more precise interpretation in terms of familiar psychological characteristics? It cannot be identified with high general intelligence ; for the items scored are by no means restricted to cognitive traits : they cover emotional and temperamental traits as well ; and in some of the categories dullards may easily gain high scores. Moreover, the factor-saturations for the two intelligence tests are comparatively low, less than 0·300.

The highest saturations (over 0·8) are for Total Number of Responses, Number of Unusual Details, and Number of Cultural Responses. These figures, and a closer study of the other saturations, both high and low, suggest that the factor can probably best be interpreted as a general factor of *fluency of productive association.*[2]

This is, after all, not an unexpected result. Recent factor studies regard fluency of productive association as the essential factor entering into what, in traditional or popular language, is termed imagination (cf. 26, pp. 180f.) ; and it may be remembered that earlier workers used the inkblot test as a test of what (in terms of the old faculty psychology) was designated ' imagination.' It is so classified by Whipple ; and Binet, Dearborn, Sharp, and Parsons all considered that its most marked characteristic was to display ' imaginative fertility.'

The category which gives the highest saturation of all is ' Total Number of Responses ' ; but, although the number of responses given by subjects with a high measurement for this factor is large, nevertheless (as the other saturations show) they do not consist of the easiest, most obvious, or most automatic types of response, such as are given by subjects possessing a ready reproductive association (e.g., responses relating to animals, anatomical resemblances, or mere ' form ' responses). In the main they include a wide variety of imaginative, emotional, or cultural responses.

[1] The fact that some of the items have been reversed does not really invalidate this statement. Thus, as ordinarily scored, ' average time ' denotes slow response ; when reversed it denotes quick response. Similarly, what Beck scores as a large number of ' popular ' or stereotyped responses, other workers would score in the reverse direction as a small number of ' original ' responses, and so on.

It should be noted, however, that many scores are expressed as percentages of the total number of responses. Now we cannot expect all types of response to be equally available whether the total number of responses is small or large (for example, as Burt points out in dealing with associative responses, " the proportion of monosyllabic words to total number of words in the sample could hardly remain the same, when the sample is small and when it is large "). Hence the percentage-scores having low saturations must consist very largely of those characteristics that cannot increase at the same rate as the total number.

[2] It may perhaps be noted that other research students in the Department working with the Rorschach tests applied to English subjects, both adults and children, have also found a general factor of ' fluency of response,' e.g., A. Petrie and S. Cox.

Rorschach's account of the ' dilated ' as contrasted with the ' coarted ' type might seem in some degree to turn on the same type of mental process : but this is not altogether true. The usual measure of the ' dilated ' type is a large score for ' the sum of colour responses *plus* the sum of movement responses ' (cf. 20, p. 267) ; but these two categories do not, as a matter of fact, yield high saturations for this factor. Further, the ' coarted ' or ' constricted ' type is characterized, not so much by low fluency, as by " rigid intellectual control."

There seems to be a somewhat closer resemblance between this factor and the characteristics attributed by Jaensch to his ' eidetic ' type. In his discussion of eidetic imagery he suggests that " certain selective principles operate on the higher levels of perception," and that what is involuntarily selected from the total number of stimuli presented and the associations they arouse " depends to a great extent on purely subjective factors such as meaning, values, conscious and unconscious wishes, interests, and previous experience." His two main types, the ' integrates ' and ' disintegrates,' reflect the presence or absence of these selective tendencies. The differences in the products of their thinking are in fact due to the way they select. Thus the ' integrates ' are the ' flexible thinkers ' ; they have ' rich associations,' and ' see meaningful wholes.' The ' disintegrates ' are more objective ; subjective valuation and experience have little effect on their perceptions ; they are ' inflexible and analytic.' It would seem therefore that the subjects who gain high factor measurements for ' fluency ' would be in many ways analogous to what Jaensch calls the ' integrate ' type.[1]

(ii) The second factor contributes as much as 19 per cent. to the total variance. It is a bipolar factor ; and the categories which have large positive or negative saturations are for the most part categories which relate to *cognitive* characteristics. Those obtaining high positive saturations are Klopfer's Form Level Score (which was designed to " get an objective scoring for the intellectual functioning " (19)), Percentage of Good Form, and Low Percentage of Form generally. Both the intelligence tests also yield high positive saturations. High negative saturations are given by Low Modified *Z*-score, Low Beck *Z*-score, Number of Bad Form Responses, Low Percentage of Responses to Whole Blot, and Number of Animal Responses. All the items involving ' details ' (e.g., *D*%, *Dd*, *AD*, and ' Human Details ') have fairly large negative saturations for this factor. It would appear therefore that this factor represents what Burt has called the ' *generalizing*,' ' synthetic,' or ' integrative ' type of attention [2] as contrasted with the ' *particularizing* ' or ' analytic ' type.

The former type of mind is apt to be more abstract and intellectual, if not always more intelligent ; the latter to concentrate rather on practical or isolated detail.[3] The contrast is in some respects analogous to what Rorschach, following Meumann, calls the *Wahrnehmungstypus* or *Erfassungstypus* (he defines it as ' mode of perceptual approach ') ; but the characteristics emphasized by the factor-saturations are by no means the same as those enumerated by Rorschach.

(iii) The third factor contributes only 7 per cent. to the total variance. The categories having high positive or negative saturations are to a large extent those commonly alleged to be indicative of the affective or *emotional* aspects of the subject's personality.

[1] Jaensch, E. R. (1930). *Eidetic Imagery*, pp. 92f.

[2] It should perhaps be pointed out that the word ' attention ' is here used in the sense adopted by Stout and McDougall, namely, as the mode of consciousness essentially involved in acts of apperception (cf. Stout, G. F., *Manual of Psychology*, chapter on ' Attention,' esp. pp. 147f. of 3rd ed.). His account of the ' generalizing ' and ' particularizing ' types of attention is largely based on Meumann's distinction between *fixierende oder konzentrative Aufmerksamkeit* and *fluktuierende oder distributive Aufmerksamkeit* (*Experimentelle Pädagogik*, 1907, I, pp. 500f.). Earlier writers had distinguished three main forms of ' imagination ' : an ' abstracting ' type, which seizes on isolated details ; a ' determining ' type, which fills out given details to make a determinate whole ; and a ' combining ' type, which creatively combines details not given in conjunction (Volkmann, *Lehrbuch der Psychologie*, I, pp. 470f. ; Sully, *The Human Mind*, I, p. 364). Meumann argues that these are properties of apperception or, as he prefers to say, attention (*op. cit.*, I, pp. 241f.).

[3] This is also a contrast noted by Binet. It appears again in his account of his two replies in the tests of ' describing an object ' and ' describing a picture ' (1, 3 : an accessible English account will be found in Myers' *Introduction to Experimental Psychology*, pp. 124f.). Myers himself, it may be added, later considered Binet's distinction between " the objective or practical type and the subjective or reflective type " of special importance in vocational guidance ; Burt stresses its possible importance in educational guidance (selecting pupils for grammar schools as contrasted with technical or modern schools). Miss Sharp's main conclusion consisted in a confirmation of this twofold classification ; and it reappears in Bartlett (7, p. 256).

Burt, in attempting to assess emotional tendencies from the Rorschach test results, follows the usual principles described by British psychologists [1] in discussing the influence of feeling on imagination and apperception ; and lays stress chiefly on " the concrete contents of the interpretation." " These," he says, " often reveal the dominant emotional characteristics of the more active appercipient interests : in an apprehensive mood the child is more likely to see ghosts, skulls, or corpses ; in a cheerful mood, to see fairies, toys, and Christmas trees. What may be called ' hedonic ' and ' valuational ' judgments are especially suggestive." Rorschach and his followers lay more stress on certain relatively abstract categories ; their inferences are based on the allusions to ' colour,' ' movement,' ' shading,' ' white spaces,' and the like ; and for the rest, they mainly rely on interpreting the ideas and images suggested as quasi-psychoanalytic symbols—the method so often stigmatized as the ' dream-book mode of interpretation.'

In Table IV A, it will be seen, the two largest positive saturations relate to the number of responses depending on Colour ($C\%$ and $C + CF$) ; on the other hand, the number of responses depending on Movement ($M\%$) yields a negative saturation, though the coefficient is not large.[2] So far this might suggest that the factor represents Rorschach's contrast between what he calls the ' extratensive ' and the ' introversive ' types. Bryn gives a tabular list of the characteristics supposed to distinguish these two types (23, p. 126) ; the chief characteristic of the ' extratensive ' is said to be their ' unstable emotional life.' However, the highest positive saturation of all relates to ' clouds ' ; and there are fairly large positive saturations for items usually supposed to be characteristic of vague, introverted,[3] and autistic thinking (e.g., vistas, landscapes, abstract responses, etc.). This type of thinking is, of course, especially characteristic of the neurotic. And much the same conjunction of categories is reported by Guirdham, who found that a marked tendency to give Colour Responses often accompanies responses of the kind called ' Diffused Shading,' e.g., references to vague and relatively formless objects like ' clouds ' or ' mist.' [4] Diethelm also noted a close relation between Colour Responses and Cloud Responses under the influence of adrenalin.[5]

All these considerations combine to suggest that this third factor tends chiefly to contrast subjects with *neurotic* and with *non-neurotic* tendencies respectively ; and this inference is, as we shall see, confirmed by external evidence.

Correlations between Factor-measurements and Independent Ratings. With a view to obtaining a more objective check on the interpretation of the various factors, approximate factor-measurements were calculated for the bipolar analysis. The

[1] Among British psychologists the writer who first laid stress on " the influence of emotional elements on imaginative combinations " was Bain : " vanity sets up pictures of admiring assemblies ; anger, once roused, resuscitates objects in harmony with it ; the emotion of terror gives a character to all the ideas formed under its influence : ghosts and hobgoblins fill the imagination of the superstitious " (*Senses and Intellect*, pp. 614f.) ; indeed, in *Mental Science*, p. 175, he proposes to " restrict the term ' imagination ' to the productive process as differentiated by the presence of emotion." Sully similarly states " the presence of feeling gives a particular direction to the imaginative process : every feeling tends to reinstate those images that are congruous with it " (*op. cit.*, I, p. 378).

[2] Rorschach himself described the interpretation of ' movement responses ' as " the thorniest problem in the entire experiment " (9, *Eng. Trans.*, p. 216). Earlier workers on apperception had contrasted ' colour,' not with ' movement,' but with ' form ' : Rorschach, chiefly it would seem for *a priori* reasons, proposed to substitute ' movement.' Burt, however, has given a good deal of evidence for maintaining that the change is not really justified (cf. 15, pp. 290f.) : " the negative correlations are between interest in form and colour, not movement and colour." This is borne out by my own results. In the correlation table (27, appendix), it will be seen, there are small *positive* correlations between colour and movement, but comparatively high *negative* correlations between colour and form. Similarly in Table V most of the Form responses have negative saturations, and ' Few Form Responses ' has a positive saturation. The highest correlation of $M\%$ is with the matrix test ($+ 0{\cdot}70$) ; and its correlation with the group test of intelligence is also one of the largest ($+ 0{\cdot}40$). This agrees with the view of many Rorschach workers that Movement Responses are indicative of intelligence.

[3] It should be noted that Rorschach's ' introversive ' type is not (as many writers appear to suppose) altogether the same as the more familiar ' introverted ' or ' inhibited ' type : the latter in many ways corresponds more to his ' coarted ' or ' constricted ' type.

[4] Cf. Guirdham, A. (1936). ' The diagnosis of depression by the Rorschach test.' *Brit. J. Med. Psych.*, XVI, pp. 130–145, and refs.

[5] Diethelm, O. (1934). ' Personality Content in Relation to Graphology and the Rorschach Test.' *Proc. Ass. Res. Nerv. Ment. Dis.*, XIV, pp. 278–286.

original scores were reduced to standard measure, and then weighted with the saturations for the three factors obtained by the bipolar method.

This device for calculating factor-measurements is exceedingly convenient in practice. But it perhaps needs a word or two in justification. As a check on its validity, more accurate weights were also calculated by Ledermann's formula, namely : $FR^{-1} = (F'U^{-2}F + I)^{-1}F'U^{-2}$, where $R =$ the correlation matrix, $F =$ the matrix of factor-saturations, $U =$ the diagonal matrix of specific factor-saturations.[1] It was found that in their relative size they closely resembled that of the corresponding factor-saturations : in other words, the post-multiplication of F' by R^{-1} or its equivalent does not greatly affect the final result. The rank correlations between the two sets of figures are for the general factor ·82, for the first bipolar factor ·91, and for the second bipolar factor ·90. The correlations between the factor-saturations and the factor-weights are surprisingly high, and suggest that comparatively little loss of accuracy results from the substitution of saturations for factor-weights.

The approximate factor-measurements thus obtained were correlated (*a*) with the results of the intelligence tests and Cattell's test of fluency, (*b*) with the average of the judges' ratings for certain personality traits, and (*c*) with the personality ratings as estimated by Burt's procedure from the content of the Rorschach responses.[2] The correlations are shown in Table V. Only traits which provide at least one saturation over 0·30 are here included. It will be seen from the various figures that the conclusions suggested by the external criteria fully confirm those based on internal evidence.

TABLE V. CORRELATIONS BETWEEN FACTOR-MEASUREMENTS AND (1) TESTS, (2 A) JUDGES' ASSESSMENTS, (2 B) ASSESSMENTS BASED ON CONTENT OF RORSCHACH RESPONSES

TRAIT	FACTOR I		FACTOR II		FACTOR III	
1. *Tests*						
Intelligence (Matrix) ..	−·132		·507		−·031	
Intelligence (No. 33) ..	·032		·456		·002	
Fluency (Cattell)	·485		·041		·194	
	A	B	A	B	A	B
2. *Assessments*	*Judges*	*Content*	*Judges*	*Content*	*Judges*	*Content*
Intelligence	·196	·180	·502	·585	−·004	−·021
Perception of Relations ..	—	·096	—	·708	—	·082
Imagination	·530	·393	·146	·202	·092	·013
General Emotionality ..	·252	·371	·156	·186	·112	·188
Extraversion-Introversion ..	·149	·307	·169	·197	−·092	−·113
Neurotic Tendencies ..	·171	·242	−·212	−·319	·685	·702

For the first factor the highest coefficient furnished by the tests is the correlation with the test of fluency, and the highest coefficient furnished by the judges' ratings and the content ratings is the correlation with imagination : general emotionality and extraversion-introversion also yield fairly large figures. For the second factor the only large correlations are those with the two intelligence tests, the two ratings for intelligence, and the content rating for perception of relations. Nearly all the other correlations are statistically non-significant. The third factor yields a significant correlation with one trait only, namely, neurotic tendencies, as assessed both by the judges' rating and the assessment made by the investigator.

[1] Ledermann, W. (1939). 'A Shortened Method of Estimating Mental Factors by Regression.' *Psychometrika*, IV, pp. 109–116. Cf. Burt, C. (1937). 'Correlations between Persons.' *Brit. J. Psych.*, XXVIII, p. 85.

[2] For neuroticism I also attempted a twofold classification, based on my observations of the subjects during the course of the tests and interviews. The correlations derived for this classification are therefore biserial coefficients, and probably less trustworthy than the other figures.

2. *Group Factor Analysis.* A supplementary analysis was also carried out by Burt's Group Factor Method (16, pp. 477f.). The original matrix of correlations was divided into three submatrices in accordance with the lines of division suggested by the bipolar analysis ; and a basic factor and three group factors were then extracted. The saturations are shown in Table IV B.

(i) The basic factor (*F*) places the categories in much the same order as did the general factor in the bipolar analysis, although the saturations have a wider range. The correlation between the two columns of saturations is 0·89. Hence it seems reasonable to regard the basic factor as representing fluency of association. It may perhaps be identified with Hargreaves' group factor of ' fluency ' (derived from inkblot and other tests) : he defines it as '*quantity* of imaginative production,' and considers that it confirms " the discovery of an Imagination factor by the authors of *A Study in Vocational Guidance.*"[1]

It may be seen that a few of the categories have small negative saturations for this factor, namely, Good Form ($F + \%$), Klopfer's Form Level Score, Form Responses generally (*F*), Movement (*M*%), Ratio of Coloured to Uncoloured Responses, and low Modified *Z*-score, as well as the Matrix test of intelligence. This probably implies that the large saturations assigned to the broad group factor (*G*), which covers these traits, really include saturations due to a narrower and more specialized supplementary factor, which, owing to the small size of the sample, cannot be significantly distinguished.

(ii) Of the three group factors the first (G) includes those traits that have positive saturations for the first bipolar factor and negative for the second. The persons belonging to this group tend naturally to generalize (they make free use of broad intellectual concepts), and are free from any marked neurotic tendencies. They thus show stable and well-organized personalities. The factor itself is perhaps best interpreted as a factor for integrated thinking, closely allied to intelligence.

Categories which are generally recognized as denoting intelligence (Klopfer's *FL*, $F + \%$, etc.), as well as Movement (which Hertz believes to indicate intelligence among children), have high saturations for this factor ; and this interpretation seems confirmed by the saturations for the two intelligence tests. The fact that the test saturations are not higher still, however, appears to imply that, with the usual scoring-categories, it is scarcely possible to isolate from the Rorschach test a clear and reliable estimate of intelligence as ordinarily understood. As we have already seen, with a larger sample, the traits covered by this factor would probably further subdivide into two subgroups.[2]

(iii) The second of the group factors (*E*) includes categories, like Colour and Few Form Responses, which have positive saturations for both the first and the second bipolar factors. According to the orthodox interpretations, it should therefore denote a factor of emotionality. Except for ' Clothing,' which has a comparatively low saturation, the characteristics having high saturations in the group factor column—characteristics which include not only ' Colour ' and ' Few Form ' Responses, but Abstractions, Landscapes, Vistas, ' Surface Grey '—are highly suggestive of the type which Burt terms the ' unstable introvert,' a type whose emotionality is of the vague, dreamy, reflective kind.[3] Burt, for example, points out that, with

[1] *Op. cit. sup.*, p. 24 ; cf. Burt *et al.*, *A Study in Vocational Guidance*, pp. 43–46, and 26, p. 181. Burt and other writers interpreted the underlying ability in terms of a positive endowment : this is implied by such terms as ' active imagination,' ' activity ' of intelligence as distinguished from ' level ' (Binet), ' cleverness ' (Garnett), and the contrast drawn between ' imagination ' and ' fancy ' (Coleridge and Wordsworth, cited by Burt). Hargreaves, however, finds nothing corresponding to this positive conception, and believes fluency of association to depend essentially upon lack of inhibition : he compares it to the ' flight of ideas ' in intoxicated and manic states. This, it may be remarked, comes near to Rorschach's account of ' dilation ' as contrasted with ' coarctation.'

[2] It will be noted that the Rorschach test, as ordinarily scored, gives little or no information about special aptitudes. Some light is thrown on these by the scores derived from an analysis of contents. Miss Hillman, in an unpublished study with children, considered that verbal and practical ability, as well as productive and reproductive imagery (the latter including imagery and immediate and delayed memory), might all be tentatively assessed, but with no very high reliability, except perhaps in the case of productive imagination, " which the inkblot material was originally employed to test."

[3] Cf. Burt, C. (1917). ' The Unstable Child,' *Child Study*, X, p. 70 (Section on ' Association ').

picture tests, the ' unstable introvert ' prefers " landscapes to portraits, scenes from dream-land rather than from reality " (15). These and other preferences suggest a close parallel to the kind of fanciful pictures which persons of this type tend to see in the inkblots.

(iv) The last of the group factors (*P*) covers all those traits that showed negative saturations for the first bipolar factor. The subjects with high factor-measurements for this factor show a tendency towards ' particularization,' in the sense described above : their responses are determined, almost naïvely, by non-constructive mechanical association. This factor therefore seems to indicate a dominance of processes akin to what Burt calls ' reproductive association ' and older writers ' passive imagination.'

The arrangement of positive and negative saturations in the second bipolar factor suggests that the traits entering into this group factor should also be subdivided. They appear to imply two distinguishable subtypes, namely, a more objective or practical type, who interpret quite impersonally the objective of concrete details which they see in the blots, and a mildly neurotic type, who interpret the details in a more personal or subjective way. Burt lays considerable stress on this latter tendency, and states that it " appears to occur much more frequently (*a*) in the feminine sex and (*b*) those persons of either sex who are of a psychoneurotic type and would often be described as suffering from ' inferiority ' or ' superiority ' complexes or even mild paranoid tendencies. Alike in tests of free association, in the interpretation of pictures, and in responses to ambiguous pictures and meaningless spot patterns, inkblots, and the like, they evince a marked tendency to give *individual* associations and interpret things in terms of their own *personal* experiences : more than half their replies are characterized by reminiscence or self-reference." The distinction appeared in Binet's experiments, and was most clearly marked in his tests of description. In his account of the results obtained from his two daughters, Marguerite (M) and Armande (A), " M," we are told, " writes much more frequently of her own past experiences than her sister : the difference is not so much due to memory as to interest. So, too, the words used by M refer more to her own person." [1]

However, with the present figures this distinction cannot be pressed. Within this set of traits the second bipolar factor shows only two large positive and two or three large negative saturations ; the rest are decidedly small. It is therefore possible that these figures merely point to a certain degree of overlapping, or possibly to a couple of small special factors linking no more than two or three items in each case.

After the foregoing factors had been extracted, there still remained a few large residuals suggestive of a definite overlapping between certain categories. The more conspicuous suggest relatively specific correlations between ' Percentage of Colour Responses ' and (*a*) ' Colour *plus* Colour-and-Form Responses ' and (*b*) ' Few Whole Responses,' and again between ' Colour *plus* Colour Form Responses ' and ' Modified *Z*-score.' [2]

The foregoing factor analyses help us to see far more clearly what it is precisely that we may to some extent infer from Rorschach scores. Although it is claimed that the Rorschach test covers the whole personality, it is obvious that it fails to give any clear information about many essential features. Indeed, it would seem that assessments based on it, particularly when scored in the usual way, are bound to be somewhat limited in value, first, because no single test and no single type of test-material can really do justice to all aspects of the personality ; secondly, because the conventional scoring has evidently been appreciably biased by the special interests of those who have developed the test. Rorschach himself and most of the more enthusiastic champions of his test have been chiefly interested in the

[1] Myers' summary, *Introduction to Experimental Psychology*, pp. 122–123. Some of the distinctions here drawn, it may be noted, are analogous to those observed by Bullough and others in their studies of artistic appreciation. Burt believes that, to a large extent, responses made in interpreting inkblots may be classified in much the same way as responses made in interpreting pictures, colours, and the like (cf. *How the Mind Works*, chapter on ' The Psychology of Art,' pp. 280f.). Miss Hillman, however, found that the correlations obtained, though usually positive, were seldom high : the largest correlations were obtained for the ' objective ' and the ' associative ' types. This seems to be in keeping with the inferences drawn above.

[2] In my thesis approximate saturations have been inserted to deal with the overlapping cases. With a larger group they could probably have been established more clearly by rotating the bipolar factors to fit the group factor pattern. Professor Burt, however, has suggested that an even more instructive form of rotation might be secured if the group factor pattern used as a guide was based upon the general structural pattern which has gradually emerged from the accumulated results of factor researches in the past. This, however, would seem to entail a larger number of independent assessments for the chief cognitive and orectic factors.

psychiatric field : hence their scoring has largely been concerned with the diagnosis of those forms of defective intelligence that are met with in the feeble-minded or the mentally diseased, and of the milder types of emotional instability that characterize the psychoneurotic and the psychopathic.

The factor analysis shows that *some* of the interpretations are undoubtedly justifiable, even if their reliability and validity are not very great. Thus the kinds of response said to be distinctive of high intelligence are, on the whole, accurately assigned. On the other hand, many of the assignments proved to be ill-founded or even fanciful. The suggestion that '*M*%' denotes 'imagination' is plainly negatived by its low saturation for factor I (0·097). The suggestion that *FY* denotes "lack of activity and listlessness" seems negatived by its high saturation for factor I. The statement that "Total *R*" is "the best indication of intelligence" seems contradicted by its negative saturation for factor II. The outstanding importance attached to *C*%, *M*%, and their sum and their ratio, is hardly warranted by the figures. And the reader who cares to compare the saturations for the several categories with their orthodox interpretations as given in the Rorschach guides will have little difficulty in discovering other instances where the usual diagnostic inferences can claim no measure of verification. Indeed, the large number of low saturations scattered throughout the table seems to suggest that many of the scoring-categories fail to confirm each other in the way that is claimed.

On the other hand, assessments based on analysis of contents seem far more promising. It is possible that a more eclectic type of scoring, based both on content and on the more suggestive of the formal characteristics, might, if subjected to a factorial study, render the test still more valuable for practical work. No doubt further studies on a much wider scale are required to substantiate the conclusions I have drawn. But the results so far obtained seem plainly to demonstrate the value of correlational studies along the lines here described.

Introspective Evidence. It is a sound principle that inferences as to the nature both of tests and of factors should be supported, not only by internal statistical evidence and by external observation, but also by introspective data, elicited from the subjects tested, and explaining the processes by which their overt responses were reached. Many Rorschach workers, after administering the test itself, conclude with a brief supplementary 'enquiry.' But this is usually directed to features in the blot, not to the mental processes of the subject. Its object is "to ascertain what part of the blot suggested the response, and whether it was determined by colour, shading, etc." : "questions should be asked only to clarify the scoring : enquiry into most responses is uninstructive and superfluous" (20, pp. 97-98). In research this restriction seems an unwise policy ; and, even in clinical work, where time is limited, it is not without serious risks.[1]

"Earlier investigations," it has been said, "on problems of this kind made the introspections the central feature of the experiment. In using indeterminate material (inkblots, spot-patterns, incomplete or ambiguous puzzle pictures, and the familiar alternating diagrams) the primary object of the investigators was to discover how, during the attention process, a relatively meaningless stimulus acquires meaning for the percipient. Trained most probably on Titchener's four-volume *Manual*, they would have learnt, from his very first experiment on attention (reporting on what is seen in puzzle pictures), that 'the appeal lies to introspection : . . . here, as elsewhere, the verdict of introspection must outweigh all preconceived opinions.'[2] Since then, however, owing largely to the reactionary criticisms of the behaviourist school, introspection has fallen into neglect. Too often the most characteristic responses are explained forthwith in terms of a small stock of alleged mechanisms, of which the percipient is supposed to be unconscious, when, in point of fact, he himself may be fully aware of the real reasons for his replies. Most frequently, of course, conscious and

[1] Miss Hillman, for example, relates how a psychiatrist at a child guidance clinic, having applied the Rorschach test to a boy of thirteen, reported that "the bizarre responses (illustrated in detail) provided strong evidence for a diagnosis of latent schizophrenia" ; whereas a subsequent introspective study of the child's mental processes showed that "seven out of the eight 'bizarre responses' were words that the child had been writing out over and over again in a spelling-exercise given that morning in school." In other instances memories of quite recent incidents, especially at the cinema, "sufficiently explained a number of the children's more peculiar replies, which would otherwise have been quite inexplicable, and were dropped entirely when the test was repeated."

[2] Titchener, E. B. *Experimental Psychology : A Manual of Laboratory Practice.* I, Qualitative, Students' Manual, p. 109.

unconscious processes both play their part. But the former are nowadays too rarely investigated in attempts at validating and interpreting tests" (10).

Many of the responses given by my own subjects could only be interpreted correctly by one who had already ascertained the particular features in their cultural background which so often prompted their replies[1] (cf. 27, pp. 33f., 93). Since the majority were untrained in introspection, the evidence available from this source was somewhat limited. Nor is a statistical journal the appropriate place for a detailed discussion of introspective studies. It may, however, be suggested that in future investigations the following points could usefully be borne in mind when investigating some of the questions raised by the correlational data and the factorial results.

As Titchener says, "introspection knows nothing of faculties." Hence, so long as we think of factors as abilities, we may gain little help. "What introspection knows are processes." Now, however meaningless the stimulus, the mental processes it arouses will inevitably be affected by "the tendency of everything psychical to secure meaning."[2] And "apperception," as Stout reminds us, "is the process whereby a presentation acquires significance for thought," or, in a word, "secures meaning."[3] Introspection thus seems essential if we are to obtain a first-hand knowledge of the various forms this process may take. With the blots, for example, some subjects will explain how the "meaning" nearly always presents itself immediately, after the briefest glance; others report that it only comes after an appreciable interval spent on what Bartlett would call "rummaging about," "searching among ancient memories now half forgotten." We have therefore to distinguish between factors operating in and during the apperceptive process itself, and those that operate outside it or before it; and between those that are conscious, and take time to evolve, and those that are more or less unconscious, and may consequently operate almost at once. Whether clearly conscious or not, all such processes will have a threefold aspect—cognitive, affective, and conative. The cognitive report, for example, may show that the subject in question nearly always finds the *name* comes first, and he then develops the implications of the name; other subjects will say that they first picture some related incident or object, and experience considerable difficulty in finding suitable words or names. But what suggests the names, the incidents, or the objects? On the affective side, an accompanying hedonic tone seems always to be present, though it may not be mentioned until an introspection is requested. If, as we are told, attention to dark shading commonly denotes an unpleasantly toned 'anxiety,' and attention to light shading a pleasantly toned 'euphoria,' introspection should, with trained psychological students at least, be able to confirm it. On the conative side, lack of fluency may or may not be due to inhibitions, or to the fact that the subject has to make what Bartlett calls an "effort after meaning." Free association and other analytic techniques may help to disentangle the deeper and less conscious influences. It may, I think, be fairly predicted that even the simplest experiments along these lines will often throw more light on the factors actually at work than either speculative interpretations or a mass of unexplained statistical data. Only, therefore, when we fully understand the *general* psychology of the Rorschach test shall we be able to use it scientifically for individual assessments (10).

[1] This, I venture to suggest, would seem to imply that, in child guidance work, it is highly inadvisable to accept interpretations of children's replies, put forward by psychologists or psychiatrists who are often quite unfamiliar with the daily life of the children in question and may themselves have even grown up in another country. It certainly confirms the criticisms so often made by teachers on reading reports by psychologists of German origin on English schoolchildren brought up in a London slum.

[2] Ewald, W. *Prinzipien der Denkpsychologie*, p. 619.

[3] *Analytic Psychology*, II, p. 110.

V. SUMMARY AND CONCLUSIONS

1. A hundred Indian students were given the individual Rorschach test, two intelligence tests, and Cattell's test of fluency. The Rorschach test was scored, first, by Beck's standard method, and, secondly, by Burt's method of rating the contents of the responses in accordance with a schedule of personality traits, such as is commonly used in individual psychology. Ratings for the various traits were also obtained from independent judges, who were close acquaintances of the subjects tested ; and these were used as validating criteria.

2. Correlations were calculated between the judges' assessments and (*a*) the individual Rorschach categories (classified according to Beck's scheme), and (*b*) the ratings based on an analysis of content. With few exceptions, the Rorschach categories, as scored by the orthodox procedure, gave figures that were statistically non-significant. On the other hand, the method of rating by content furnished consistently high coefficients.

3. Correlations between the 36 Rorschach categories were factorized by (*a*) simple summation and (*b*) the group factor method. The interpretation of the factors obtained was based partly on the internal evidence of the saturations themselves, partly on the external evidence supplied by independent assessments, and partly on introspective evidence obtained during the supplementary enquiry. The bipolar analysis indicated, first, a general factor of fluency, and, secondly, two bipolar factors concerned with cognitive and with emotional traits respectively. The first of the two bipolar factors appeared to contrast responses of a synthetic or generalizing type with those of an analytic or particularizing type ; the second appeared to contrast neurotic with non-neurotic tendencies. The group factor analysis suggested a basic factor of fluency, and three group factors which might be interpreted as representing very approximately general intelligence (an integrative or generalizing factor), general emotionality (with a marked bias towards introversion), and a relatively mechanical tendency towards the associative reproduction of particulars.

4. To provide a check on the inferences drawn from the factor-saturations, approximate factor-measurements were calculated for each factor ; and these were correlated with (*a*) the results of independent tests, (*b*) judges' ratings for the more relevant traits, and (*c*) assessments based on an analysis of the content of the responses. On the whole the correlations appeared to corroborate the tentative identifications of the several factors.

5. The results of the factor analysis confirmed several of the orthodox interpretations regularly suggested for some of the commoner types of response, particularly for those alleged to be indicative of intelligence and of neurotic tendencies. Many of the more formal categories, however, appeared to have little or no general psychological significance ; and the inferences commonly drawn from responses referring to colour, movement, and the like seem to rest more on preconceived hypotheses than on verifiable evidence. It is concluded that an analysis of content, based on generally recognized psychological principles, and expressed in terms of generally recognized psychological factors, would be far more valuable for practical purposes.

REFERENCES

1. Binet, A., and Henri, V. (1895). ' La psychologie individuelle.' *L'Année Psychologique*, II, 411–465.
2. Dearborn, G. (1897). ' Blots of ink in experimental psychology.' *Psych. Rev.*, IV, 390–391.
3. Binet, A. (1903). *L'étude experimentale de l'intelligence.* Paris : Alcan.
4. Whipple, G. M. (1910). *Manual of Mental and Physical Tests.* Test 45. 430–435.
5. Burt, C., and Moore, R. C. (1912). ' Mental differences between the sexes.' *J. Exp. Pedag.*, I, 273–284, 355–388.

6. Pyle, W. H. (1913). *The Examination of School Children.* New York : Macmillan.
7. Bartlett, F. C. (1916). ' An experimental study of some problems of perceiving and imagining.' *Brit. J. Psych.*, VIII, 222–267.
8. Parsons, C. J. (1917). ' Children's interpretations of ink-blots.' *Brit. J. Psych.*, IX, 74–92.
9. Rorschach, H. (1921). *Psychodiagnostik : Methodik und Ergebnisse eines Wahrnehmungs-diagnostischen Experiment.* Bern : E. Bircher.
10. Burt, C. (1934). *The Rorschach Test.* Laboratory Notes, University College, London. (Revised 1945.)
11. Kerr, M. (1934). ' The Rorschach test applied to children.' *Brit. J. Psych.*, XXVII, 51–59.
12. Vernon, P. E. (1935). ' On the significance of the Rorschach test.' *Brit. J. Med. Psych.*, XV, 199–207.
13. Vernon, P. E. (1935). ' Recent work on the Rorschach test.' *J. Ment. Sci.*, LXXXI, 1–27.
14. Vernon, P. E. (1936). ' The matching method applied to investigations of personality.' *Psychol. Bull.*, XXXIII, 149–177.
15. Burt, C. (1939). ' The factor analysis of emotional traits.' *Character and Personality*, VII, 238–254, 285–299.
16. Burt, C. (1940). *The Factors of the Mind.* London : University of London Press.
17. Burt, C. (1945). ' The assessment of personality.' *Brit. J. Educ. Psych.*, XV, 107–121.
18. Beck, B. J. (1945). *Rorschach's Test.* New York : Grune and Stratton.
19. Klopfer, B., and Kelley, D. (1946). *The Rorschach Technique.* New York : World Book Company.
20. Rapaport, D. (1946). *Diagnostic Personality Testing.* Chicago : World Book Company.
21. Cattell, R. B. (1946). *The Description and Measurement of Personality.* New York : World Book Company.
22. Eysenck, H. J. (1947). *Dimensions of Personality.* London : Kegan Paul.
23. Bell, J. E. (1948). *Projective Techniques.* New York : Longmans, Green and Co.
24. Ramzy, I., and Pickard, P. (1949). ' The reliability of Rorschach scoring.' *J. Gen. Psychol.*, XL, 3–10.
25. Goodenough, F. L. (1949). *Mental Testing.* New York : Rinchart.
26. Burt, C. (1949). ' The structure of the mind : a review of the results of factor analysis.' *Brit. J. Psych.*, XIX, 176–199.
27. Sen, A. (1949). *A study of the Rorschach test.* Doctorate Thesis, University of London Library.

GROUP FACTOR ANALYSIS

By CYRIL BURT
University College, London

I. *General Problem.* II. *Non-Overlapping Group Factors :* A. *Three Group Factors.* B. *More than Three Group Factors.* C. *Two Group Factors.* D. *One Group Factor.* III. *Overlapping Group Factors.* IV. *Summary and Conclusions.*

I. GENERAL PROBLEM

Aim. In the following paper my main purpose will be to state the essential principles underlying what I have called the ' group factor method ' of factor analysis, and to describe the simplest working procedures available for the commoner types of problem that occur in actual practice. This will give me an opportunity of replying more fully to criticisms of the method raised by earlier writers, and of answering relevant questions received from recent correspondents. At the same time I shall endeavour to explain in what ways the underlying concepts resemble, and differ from, those involved in the related notions of ' simple structure ' and ' bifactor analysis.' The fact that the recent edition of Professor Thomson's book now includes an ' Addendum ' on the ' bifactor method ' has aroused a special interest among British workers ; and has led to several enquiries about its relations to previous procedures, which I shall attempt to answer so far as my own limited experience of it permits.[1]

Need for a Group Factor Method : (1) *With Physical Measurements.* In several previous contributions I have described how the notion of factorizing a correlation table into group factors originally arose (7, 8, 9, and refs.). Here I need only recapitulate the history of the concept quite briefly.

The need for some such procedure was encountered first in connexion with the analysis of physical measurements. To make the problem clear, let us begin with an imaginary instance. Suppose we have taken four representative measurements for the size of the human body (e.g., height, weight, chest breadth, and arm length) and three for the size of the head (e.g., length, breadth, and height) ; suppose, further, that we have found large positive inter-correlations between the first four traits, and large positive inter-correlations between the last three traits, but zero or non-significant cross-correlations between the first four and the last three : in such a case we should, I imagine, conclude that there were two distinct and separate growth-tendencies at work—one for the growth of the body and another for the growth of the head.

So obvious a conclusion would be drawn regardless of any thought or theory of factorial analysis : it would have been unhesitatingly inferred by Galton, Pearson, or any of their followers half a century ago. It is, however, convenient to have a more general name for

[1] *Factorial Analysis of Human Ability* (1948), pp. 341f. In this country the bifactor method seems almost unknown. The only large-scale enquiry in which it has (to my knowledge) been used is an unpublished research on ' Temperamental Factors,' recently carried out by Mr. C. J. Adcock (Ph.D. Thesis, University of London, 1937). Thurstone, in his recent work (*Multiple Factor Analysis* (1947), p. viii), complains that Holzinger and Harman " dispose of the [his] simple-structure concept in a footnote " (10, p. 102) ; but he himself dismisses the bifactor method with only a parenthetical sentence in his preface. In my own view, as I have said elsewhere (7, 9), the different methods put forward by different writers should not be regarded as rival competitors : each may have its special merit for the purpose for which it was contrived. Hence I hope it will be understood that my criticisms relate primarily to the particular kind of problem on which my fellow-workers and I have been principally engaged.

such hypothetical tendencies ; we may therefore agree to call them ' factors.' And, since they are factors that enter into limited groups of traits only, it is natural to refer to them as ' group factors.'

If, as psychologists, we are interested more particularly in the growth of the head or skull, we might go on to enquire which of the three head measurements, or better still what weighted sums of all three measurements, would give the best indication of skull-growth, and indirectly, we might suppose, of brain-growth. By means of such weights we could obtain explicit ' factor-measurements ' for our ' group factor.'

Now, when we turn to an actual table of such correlations, for example, the table for physical measurements discussed by Pearson and Macdonell in 1902,[1] the pattern discovered is not so clear cut : we find that, while the inter-correlations within the two main groups are both positive and large (averaging 0·67), the cross-correlations are by no means negligible ; though relatively small (averaging only 0·27), they are all positive and all significant. Thus, as Pearson expresses it, *both* ' like ' *and* ' unlike ' traits are positively correlated, although the correlations between ' like ' traits are much higher. Evidently two independent and non-overlapping group factors will not suffice to explain the figures actually observed.

Pearson himself, it will be remembered, suggested extracting as many uncorrelated ' factors ' (or ' index characters,' to use his own term) as there are correlated traits, and choosing for that purpose the ' principal axes ' of the correlation ellipsoid. In that case all factors except the first would necessarily be bipolar. However, on examining that portion of his table which gave the cross-correlations, it was observed that, curiously enough, the figures were almost exactly ' synclinal ' or ' hierarchical,' i.e., they virtually formed an oblong *submatrix of rank one* (see 12, Table I, p. 104). It therefore seemed uneconomical to use seven factors to explain these particular figures when one factor alone would serve. And in fact a very close fit to the whole table was obtained with three factors only, one ' general ' and two ' group ' (*ibid.*, p. 118).

(2) *With Mental Tests.* Similar correlation patterns were from time to time observed in the early results obtained for mental traits. But it was not until about 1924 that the idea of group factors became widely accepted by psychologists.[2]

In my first paper on intelligence tests (1909) I found evidence for " groups of allied tests to correlate together " : thus, after deducting the effects of general intelligence, there still remained significant or nearly significant residuals for the inter-correlations (*a*) between tests of sensory discrimination, (*b*) between tests of motor co-ordination, (*c*) between sensori-motor speed tests, and (*d*) between tests and assessments involving memory (1, pp. 144, 164, and Tables V and VI). These results were confirmed and extended in an investigation carried out with Mr. Moore on much larger groups of schoolchildren, tested by means of group tests, which included ' higher mental processes.' Independent corroboration seemed to be afforded by Simpson's work with adults.[3] Simpson applied 14 tests, representing different cognitive levels (similar to the five ' levels of mental process ' tested in my own investigation), to 37 adult males ; he also found, in addition to ' general mental ability,' ' certain relatively specialized capacities ' belonging to particular categories or groups (very similar to those I had already reported), namely, (*a*) sensory discrimination, (*b*) motor control. (*c*) memory and association, (*d*) quickness and accuracy of perception, and (*e*) selective thinking. Simpson, however, based his inferences merely on a comparison of the observed coefficients, and made no attempt to eliminate the general factor or test the residuals for significance.

Our conclusions were at first strongly criticized by Spearman and others. Not without justice, he objected, first, that the statistical techniques employed were crude and inconclusive, and secondly that the numbers tested were too small for the results to be accepted as fully significant. Judged by the intercolumnar criterion, he found that my own data gave a mean corrected correlation of over 1·00 in both investigations, and Simpson's one of 0·96. Moreover, in common with other

[1] *Biometrika*, I, 1901–02, pp. 177–207. In this enquiry the dimensions of body and head actually measured were not quite the same as in the hypothetical example I have taken in the text, since the investigators had to choose their figures from the Bertillon scheme of measurements collected by Scotland Yard from criminals.

[2] See *Report on Psychological Tests of Educable Capacity*, Board of Education, 1924. In Appendix IX, the majority of psychologists agreed in accepting, at least provisionally, a hypothesis including both a general factor and a number of group factors. A review of the chief group factors that now seem to be fairly well established will be found in *Brit. J. Educ. Psych.*, XIX, 1949, pp. 100–111, 176–199.

[3] ' Correlations of Mental Abilities,' *Columbia University Contributions to Education*, No. 53, 191

writers at that date, he doubted whether the distinction between ' like ' and ' unlike ' traits, so conspicuous in the case of physical measurements, could be expected to hold good in a field so different as that of mental testing, " unless the ' similar ' tests were so similar as to be virtually identical."

American criticisms were less emphatic. Freeman, for example, observed that " Simpson presents his results in the form of a table similar to that which Burt reported, and the order of the tests bears out the same conclusion which was drawn from Burt's " : nevertheless, he had " doubts as to the validity of the classification " (" according to mental categories "), largely because " the correlation between traits in different classes is often much closer than the correlation between traits in the same class. . . . If Spearman's two-factor theory is correct, special capacities do not fall into groups." [1]

(3) *With Educational Abilities.* Conclusive evidence for the existence of such ' group factors ' was first obtained in assessments of educational abilities. Here written tests could be employed ; and data could consequently be secured from samples large enough to render the first residuals fully significant and the second residuals definitely non-significant—at least in the earlier studies.

To estimate the correlation of each trait or test with the hypothetical factors, a method of ' simple summation ' was employed, based originally on the assumption that the ' observed average ' of a given set of correlations could be identified with the ' true average.' In dealing with residual correlations left after the effects of the first or general factor had been eliminated, two courses seemed open : " we may look either (*a*) for limited group factors, shared only by similar sets of tests, whose influence will be solely positive (the natural assumption in analysing intellectual abilities), or (*b*) for supplementary general factors, ambivalent in their nature, which will account not only for specific resemblances but also for specific antagonisms (an assumption that appears perhaps more plausible in analysing emotional and temperamental differences). Moreover, these alternative aims will, as a matter of fact, affect our conception of the general factor, and the way in which we estimate it. . . . When we come to the secondary factors, with assumption (*a*) the obvious working-procedure will be to take the significant positive correlations " (i.e., the clusters of positive residuals) " in subgroups, and factorize these just as we should complete tables." [2]

It was by this latter procedure (' Method *a* ') that, working in collaboration first with Mr. R. C. Moore and later with Miss M. Bickersteth (3), I obtained what seemed to be unquestionable evidence for the presence of special cognitive abilities in both mental and scholastic tests—abilities of a kind which at that time Spearman and others strongly denied on the ground that they resembled the old-fashioned ' faculties.' The first explicit confirmation of our results came from one of Spearman's own pupils, Miss N. Carey, who based her deductions on school marks. " With actual tests of school abilities," she writes, " Mr. Burt has found [3] a ' general educational ability,' combined with ' specific educational abilities ' (arithmetical, linguistic, literary, and manual), very similar to the combination of the ' hypothetical general factor ' with subordinate ' group factors ' that he found in his intelligence tests " ; and she points out that her correlations for school marks indicate much the same combination of three types of factor, namely, in addition to (i) the ' general '

[1] Spearman's most detailed discussion of Simpson's result is to be found in his *Abilities of Man* (6, pp. 145f.). He there applied his later ' tetrad-difference criterion ' to Simpson's correlation table (it was the only large table for which this criterion was worked out in detail) ; and concluded that the resulting graph " displays one of the most striking agreements between observation and theory " [the theory of a single common factor] " ever recorded in psychology " : in a footnote, however, he admitted that there are " some very few cases . . . more than five times the probable error " (and a still larger number, it might be added, more than *three* times the probable error). Freeman's criticisms, embodying earlier comments, will be found in his *Mental Tests* (1926), pp. 77–80.

[2] *Annual Report of the Psychologist to the L.C.C.* (1914–15), Appendix I : cf. 7, p. 306 and pp. 62–68 below. Here I should perhaps explain that the phrases used to designate the two alternative procedures have varied considerably. The full titles should be ' summational analysis into general and bipolar factors ' and ' summational analysis into general (or basic) and group factors ' : but these circumlocutions are exceedingly cumbrous. In early memoranda on working methods (quoted in 7, p. 306, and 8, p. 356) the two alternatives were simply described as ' Method *a* ' and ' Method *b* ' respectively. Later, ' Method *a* ' was called ' analysis by submatrices.' Nowadays the procedures are, as a rule, referred to as ' bipolar ' and ' group factor ' analyses, it being tacitly assumed by most British factorists that there will usually be a ' general ' or ' basic ' factor as well. The use of different designations in the numerous published and unpublished researches, as well as the need for slightly different working formulæ in different cases, has, not unnaturally, prevented non-statistical readers from noting that the essential principles have remained the same.

[3] Miss Carey is referring to results reported in the *Annual Report of the Psychologist to the L.C.C.* (1914) (briefly summarized in 3, pp. 34–37). Appendix I of the *Report* contained a short account of the derivation of the formulæ (afterwards included in *Mental and Scholastic Tests*, 2nd ed., pp. 271–275).

and (ii) the ' specific factors,' (iii) a series of ' group factors ' : (in her own data, these last comprised a verbal factor for verbal subjects, a motor factor for handwork subjects, and probably a third factor for informational subjects). To simplify the computations, however, she proposed to modify my procedure by assuming that the residuals on which any one group factor is based are approximately all the same in size. Incidentally she observes that her results are in marked conflict with the assumptions made by Spearman and Hart in examining my earlier paper, namely, that the high intercolumnar correlations obtained from tables like my own commonly demonstrate that there is *no* such clustering of high coefficients as the presence of group factors would imply. She herself actually obtains *negative* correlations between columns when these relate to tests that contain different group factors and low *g*-saturations ; and she uses this result to justify the way she has partitioned her table.[1]

In a number of later researches and theses, the group factor method was regularly employed. But for the most part the working procedures adopted were improvised to meet the special needs of each particular enquiry ; and, as the interest lay rather in the psychological results obtained than in the mathematical devices used, little trouble was taken at the time to establish them systematically on a comprehensive basis. Indeed, during the 1920s, it would have been commercially impossible in this country to publish a book containing numerous correlational formulæ and expensive tables except with the assistance of the London County Council ; and the Council, quite properly, was more concerned to publish accounts of the practical conclusions reached than discussions of statistical procedures. As a result, during and since the war, younger psychologists have tended to substitute the methods described in the more accessible American publications, particularly the ' centroid method ' and the device of ' simple structure.' Nevertheless, it has apparently proved difficult to dispense altogether with the ' general ' or ' basic ' factor, as ' simple structure ' seems to require.[2] And, partly for this reason no doubt, many requests have been recently received for a factorial method which would take such a factor into account. However, before attempting to illustrate in detail the working procedures suggested, it will be advisable to formulate a little more precisely the underlying hypothesis and the fundamental principles on which all of them are based.

The Hypothesis Involved. The early work I have just described seemed to make it more and more evident that, in all researches on individual psychology and especially in the field of cognitive and scholastic testing, a possible interpretation to be kept constantly in mind was what I have termed the ' group factor hypothesis.' The phrase is convenient, but curtailed. I use it to denote the hypothesis that the traits tested or measured can best be analysed into (i) a basic or general factor, entering into *all* the tests, and alone accounting for the cross-correlations between the several groups, and (ii) a set of group factors, either overlapping or non-overlapping, each

[1] Carey, N. (1916). *Brit. J. Psych.*, VIII, p. 180. Miss Carey's procedure departs in one or two minor respects from my own. In previous discussions both Pearson and Spearman had criticized the way I had proposed to deduce, from restricted parts of the table only, a general factor which by definition was to cover the whole table.

Pearson insisted (and I think rightly, so far as a rigorous proof is concerned) that the first step should be to show that significant residuals were left, even when the closest possible fit to the table as a whole had been obtained (i.e., to use factorial terminology, that the general factor should be calculated from the entire table by the method of least squares, or some equivalent procedure). Further, he argued (as mathematical critics of current factorial methods are still inclined to do) that the leading diagonal should contain unity, not reduced self-correlations. When we try to discover the minimum self-correlations from unrepeated tests, we run the risk of trying to deduce more independent variables than our table permits, and we further render any valid test of significance doubly difficult.

Spearman insisted that the general factor ought rather to be based on some external criterion, e.g., teachers' assessments or selected tests used as reference values ; and he considered that a true ' hierarchy ' should show *equal* decrements, not rows of coefficients diminishing in proportion. Accordingly, since Miss Carey was venturing to criticize Spearman's own conclusions, it was necessary for her to adopt his general procedure. Thus, in calculating the saturation coefficients (" the correlation of each school subject with *g* "), she used Spearman's method, which derived the ' saturation ' not from the table of observed inter-correlations, but from two special tests assumed to give pure measures of *g*. In her main argument, too, she assumed that the hierarchy would be additive, though towards the end of her paper she is inclined to adopt my own procedure (cf. pp. 175 and 182).

[2] Vernon, for example, observes that, " in spite of the fact that no Navy or Army psychologist was an ardent disciple of Spearman, . . . over and over again our test inter-correlations have yielded a general factor running through all the tests and covering 30 to 40 per cent. of the variance, together with two, three, or more group factors " (' Statistical Methods in the Selection of Navy and Army Personnel,' *J. Roy. Stat. Soc.*, VIII, 1946, p. 145).

of which is primarily limited to a specific cluster or *group* of tests (presumably measuring ' like ' or ' similar ' mental functions), and in all of which the significant saturations are essentially positive.

In this hypothesis there was nothing new except the attempt to define the relevant concepts with sufficient precision to permit of rigorous statistical verification. Some such scheme of factors is already implied in the traditional attempt to classify mental faculties or functions into comprehensive *genera* and subordinate *species*. It is still more clearly implied in Galton's doctrine that an individual's achievements in a particular sphere may be explained in terms of some ' specific aptitude ' for that sphere of work, superposed upon a ' general ability ' required in every sphere.

A Working Procedure : General Principles. Having defined our terms, our next task is to deduce the verifiable corollaries of this particular hypothesis, and see how, in any concrete case, its consequences can be tested by means of the data actually obtained. Where more than one factor was concerned in any given test, it was assumed that, whether cognitive or conative, ' general,' ' basic,' ' bipolar,' or ' group,' the factors would combine (like velocities and forces in ordinary dynamics) in accordance with the ' parallelogram law.' [1] From this it follows that, if the factors postulated are independent and uncorrelated, the correlation between any two tests, r_{xy} say, could be expressed as a product-sum of the factor-weights or ' saturations,' i.e.,

$$r_{xy} = r_{xb}r_{yb} + r_{xc}r_{yc} + \ldots,$$

where the hypothetical correlations between the tests and the factors (r_{xb}, r_{yb}, etc.) are taken as weights (see Appendix, p. 72, equation iv). This in turn implies that any observed correlation matrix is regarded as analysable into a sum of superposed ' hierarchies ' (i.e., matrices of rank one), each of which may cover either *all* the traits tested or only *some* of the traits tested (8, p. 344).

As applied to the group factor hypothesis, these assumptions lead to two convenient requirements which any working procedure should fulfil. (i) First, " so far as possible, all the *significant* residuals (on which the supplementary factors were to be based) should be positive, never negative." (ii) Secondly, " all the *non-significant* residuals should for any given test (i.e., for any given row or column of correlations where only one factor was involved) sum to zero."

Each of these conditions requires a comment.

(i) To most teachers and educationists, it would certainly appear clumsy and far-fetched to classify cognitive traits in bipolar terms, and to postulate ' abilities,' or at any rate ' factors,' which have about as many negative saturations as positive. Indeed, with a few notable exceptions, both logicians and scientists have commonly been averse from dichotomous classifications when others are available.[2] Accordingly, in researches like those I have described, it seemed simpler and more intelligible to postulate positive group factors for verbal, arithmetical, and manual tests respectively (in accordance with ' Method *a* '), instead of classifying the various performances first into ' verbal ' or ' non-verbal,' and then subdividing the ' non-verbal ' (and sometimes quite illogically the ' verbal ') into ' manual ' and ' non-manual,' and so on (as required by ' Method *b* '). No doubt, even with educational abilities the occurrence of an occasional residual with a negative sign is not of necessity inexplicable. But it should be regarded as exceptional.

(ii) If we desire to judge ' goodness of fit ' by the principle of least squares, then the closest fit will be obtained by taking a *weighted* sum of the non-significant residuals : for then our aim will be to minimize the square-sum of all the non-significant discrepancies. This entails a rather elaborate process of successive approximation. In practice, however, it appeared sufficient for most purposes to substitute the principle of ' simple (i.e., unweighted) summation,' similar to that adopted in calculating an ordinary average or a mean deviation. This yields a direct and simple routine procedure for obtaining factors of the type required : (see Appendix, pp. 72–73 below).

[1] This assumption was first put forward in attempting to analyse the ' general factors and specific factors underlying the primary emotions ' (cf. *Brit. Ass. Ann. Rep.*, 1915, pp. 694–696) : it was an obvious corollary to McDougall's view of emotions as essentially conative ' forces.' But almost immediately it became plain that, from a formal standpoint at any rate, cognitive capacities could be treated in precisely the same way.

[2] Jevons, F. B. (1900). *Principles of Science*, ch. XXX ; Joseph, H. W. B. (1931). *Introduction to Logic*, pp. 121–122.

Accordingly, let us now consider what are the most convenient procedures for factorizing a given correlation table along these lines, and how we can best test the 'goodness of fit' furnished by factors obeying the scheme described. We may take the simplest form of the hypothesis first of all, namely, that which assumes that there is no overlap between the various group factors. As we shall see, the double principle just described leads to slightly different formulæ and slightly different working methods according to the number of group factors required.

II. NON-OVERLAPPING GROUP FACTORS

Partitioning the Correlation Table. The device on which the proposed procedure depends is that of 'condensation.' After the observed table of correlations has been appropriately rearranged and partitioned, in accordance with the grouping of the tests, the various figures or 'elements' in each submatrix are summed, so as to reduce the entire matrix of observed correlations to one of lower order. Provided the groups show no overlapping, this can then be treated as an approximate matrix of rank one, except for those sums which now appear in the diagonal.[1] But, before we can proceed to the actual calculation of the group factor saturations, we must first decide how the original table is to be partitioned, in other words, how the tests or traits are to be grouped.

Thomson, in an appendix on the bifactor method, added to the last edition of his book, mentions three alternative suggestions. But one or two others may also be put forward.

(1) *Subjective Classification.*—Probably the first and most natural suggestion would be that the grouping should be made "subjectively, by considering the nature of each test, and putting together the memory tests, or tests involving number, and so on" (11, p. 341). As numerous theses show, this is a favourite method with the youthful investigator. Yet it seems highly inadvisable. It provides too easy a way of appearing to establish the very hypothesis the investigator hopes to prove. The purpose of factorization is to allow the analysis to reveal, quite automatically and objectively, what are the actual groupings. If, however, we start by *assuming* (say) that all our memory tests form one group and all our number tests another, then we shall be basing the method of calculation on the very conclusion which the method of calculation ought automatically to reveal.[2]

(2) *Coefficient of Belonging.* Holzinger prefers to calculate a 'coefficient of belonging.' This is a somewhat slow and laborious preliminary; and, as ordinarily used, it does not always lead to a unique subdivision. Most investigators seldom calculate every possible value for the crucial coefficient; and it is then all too easy to stop when the results correspond with the investigator's preconceived desires.[3]

(3) *Correlation Profiles.* A third method is to make a graph or profile of each row of correlations, grouping those tests with similar profiles. This procedure has been most strongly advocated by Tryon. Thomson states that he himself prefers to compare, not the contour of the profile as a whole, but "only the peaks of each row." The device undoubtedly throws light on the structure of the correlation table [4]; but, in many cases if not in most, it does not yield a very satisfactory

[1] Numerous short-cuts depend upon this general principle, e.g., the 'sum method' of bipolar analysis (see, for example, this *Journal*, II, i, p. 61).

[2] Of course, it is possible that the whole research may have been planned to test some particular hypothesis, involving a specified grouping of the traits. But in that case it is essential to show that the group factor matrix reached on this hypothesis is the only hypothesis that gives a satisfactory fit, or at least gives a better fit than any other hypothesis. In the simpler type of problem this can be done quite simply by showing that a bipolar analysis leads automatically to the grouping assumed.

[3] This has been the experience of a good many workers in this country, as will be seen from their researches. One of the most recent and convincing set of illustrations will be found in Wheeler's thesis on *An Analysis of Tests of Manual Dexterity* (Univ. London Library) which I hope may shortly be published.

[4] In early memoranda I suggested graphing correlations *after the table had been rearranged* (*a*) to determine how far the table was 'synclinal' ('hierarchical') or 'heteroclinal,' and (*b*) to estimate the self-correlations. The idea was borrowed from Karl Pearson, who called such graphs 'analographs.' The procedure presents an interesting picture of the relations suggested; but it was found to be too lengthy and inaccurate to form a regular part of the routine procedure.

way of partitioning the table, because the profiles and the peaks of the observed correlations may be affected both by the general factor and by the errors of sampling.

(4) *Intercolumnar Correlations.* Spearman, it will be remembered, in contrasting the older hypothesis of ' multifocal levels ' with his own ' unifocal ' hypothesis, suggested, as a crucial test, the value of the correlation between the various columns of observed correlations. " When the columns belong to [tests from] different levels, the high coefficients would always come in non-corresponding places, so that the correlation between columns is *negative* " (*Brit. J. Psych.*, V, p. 57 ; cf. 6, p. 139). With my own table, which he reprints, he finds the (uncorrected) intercolumnar correlation to be, for the four largest cases, + 0·81, for the five smallest + 0·59. He therefore infers that, contrary to my own contention, the table gives no evidence whatever for group factors. Miss Carey, however, has pointed out that, unless the amount of correlation due to the group factors is large and the amount due to the general factor comparatively small, the intercolumnar coefficients will *not* be negative, but only reduced in size. In practice therefore, particularly if there is some measure of overlapping, the criterion may often prove somewhat inconclusive.

(5) *Preliminary Bipolar Analysis.* The method which to my mind is by far the best for general purposes is to start with an ordinary factor analysis by simple summation, and then to take the positive and negative sections of the bipolar factors (provided their saturations are significant) as indicating where the lines of partition are to be drawn. This procedure is entirely automatic and objective. In certain cases a preliminary analysis of this kind may be dispensed with : when the correlations have been rearranged so as to bring clusters of high coefficients as near to the diagonal as possible, it will sometimes be found that the lines of classification are sufficiently clear without further check. But such cases are exceptional : whenever there is the smallest room for doubt, an initial bipolar analysis is in my view indispensable.

The Addition and Multiplication Procedures. In the group factor method, as compared with other forms, the chief novelty arises from the way the first or ' general ' factor has to be computed. Since it is not precisely the same as the ' general ' factor reached by ordinary ' simple summation,' it is perhaps better to give it a distinctive name. As I have stated elsewhere, the underlying aim is to " seek a general factor which will represent, not the *average* slope or plane of the correlations, but the *basic* slope " (7, p. 308). Accordingly, we may call the factor so obtained the ' basic ' factor. Our discussion will therefore be concerned chiefly with the different devices available for calculating this factor.

With the group factor hypothesis, as we have seen, the essential assumption is that all the factors except the first are limited (save for a sporadic overlap) to groups of ' like ' tests ; and that in consequence *those parts of the correlation table which relate to ' unlike ' tests will form parts of a ' hierarchy '* (matrix of rank one), since they are due solely to the single basic factor (except for occasional overlaps which we have agreed to consider later). Now, in an ideal case, where the correlations in this portion of the table are *exactly* hierarchical, it is easy to show [1] that the fundamental equation given above yields two alternative procedures for calculating the saturations, which in such a case will furnish precisely the same figures. The first requires us to *add* the row-totals ; the second to *multiply* them.

The multiplication method leads to a formula which may at first sight seem simpler ; and it forms the basis of the revised form of Holzinger's ' bifactor method.'[2] On the other hand, if I may

[1] To save space I shall not give algebraic proofs for all the formulæ implied by the procedures here described. The more important are outlined in the Appendix below, and are discussed in greater detail in the roneo'd *Laboratory Notes on Factor Analysis : Group Factor Method.* There the fuller exposition gives, for each procedure, (i) the algebraic derivation of the method ; (ii) a simple illustration, based on an artificial correlation table, exemplifying the essential principle ; (iii) an application to an empirical correlation table. From the discussion and references below, the interested reader will, I imagine, have little difficulty in supplying more details for himself, and finding appropriate examples in the literature.

[2] Holzinger, K. J. (1937). *Student Manual of Factor Analysis*, pp. 13f. As originally described, the ' bifactor method ' assumed that the common factors could be " thought of as split into two *proportional* parts, one of which is the principal factor." As a consequence, the saturation coefficients given for the secondary factors in the illustrative tables were always *exactly* proportional to the corresponding coefficients for the principal factor : cf. Holzinger, K. J. (1935), *Preliminary Reports on the Spearman-Holzinger Unitary Trait Study*, No. 5, p. 6 and table on p. 2. This conception was originally put forward by Spearman, who sometimes called it a ' bifid factor.' It is referred to by several of Spearman's students, e.g., Hargreaves and Stephenson, but was not, I think, actually adopted in any of their investigations. The earliest published reference to it is in the appendix to H. L. Hargreaves' monograph, where he discusses the alternative forms which Spearman's theory

trust my own experience, what I have called the addition method appears preferable in actual practice. In my original memorandum two reasons were briefly mentioned for this preference. First, where the observed coefficients (owing it may be to large fluctuations due to sampling) depart appreciably from a strict hierarchical arrangement, there the multiplication method is apt to exaggerate the effect of these departures : in certain cases it may even give saturations over unity ; in others, where one of the sums to be multiplied is nearly zero, it may give saturations that are too near to zero. Secondly, the multiplication method does not lend itself to the usual type of check. In view of the questions that have recently been raised, these points perhaps deserve exemplification at greater length.

A. THREE GROUP FACTORS

Example. May I begin with what is at once the simplest and the commonest type of case, namely, that in which there are three group factors ? Most of the worked examples given to illustrate this or allied procedures are based on tables of artificial correlations which give exact results. In such cases the figures obtained with Holzinger's method and my own are likely to be identical. To illustrate the differences it is therefore desirable to choose a table that deviates to some extent from the exact hierarchical figures. To keep the working as simple as possible, let us take a table involving two tests only in each group (Table I). As usual, the chief difficulty arises, not over the calculation of the group factors, but over the extraction of the basic factor. I shall therefore omit the square submatrices affected by the group factors.

TABLE I. CALCULATION OF SATURATIONS FOR THE BASIC FACTOR
Matrix of Observed Correlations, Partitioned and Condensed

Tests	1	2	Sum	3	4	Sum	5	6	Sum
1	—	—	—	·40	·26	·66	·44	·30	·74
2	—	—	—	·55	·38	·93	·05	·02	·07
Sum ...	—	—	—	·95	·64	1·59(*A*)	·49	·32	·81(*B*)
3	·40	·55	·95	—	—	—	·18	·10	·28
4	·26	·38	·64	—	—	—	·11	·09	·20
Sum ...	·66	·93	1·59(*A*)	—	—	—	·29	·19	·48(*C*)
5	·44	·05	·49	·18	·11	·29	—	—	—
6	·30	·02	·32	·10	·09	·19	—	—	—
Sum ...	·74	·07	·81(*B*)	·28	·20	·48(*C*)	—	—	—
Grand Total	1·40	1·00	2·40	1·23	·84	2·07	·78	·51	1·29
Divisors ...	1·4652			2·1325			2·6087		
Sat. (Add.)	·9555	·6825	1·6380	·5768	·3939	·9707	·2990	·1955	·4945
Sat. (Mult.)	1·0087	·3682	—	·5731	·3975	—	·2989	·1956	—

With both the addition and the multiplication procedure the work is considerably facilitated if the sub-totals for the row-sections and column-sections are calculated at the outset, and entered in the table of observed correlations as shown above ; the totals for the submatrices and for the

of a general factor might take : among the conditions under which " a general factor can be compound, and so act as a single general factor," one is that in which " two (or more) factors enter into every test in exactly the *same proportion* " (*Brit. J. Mon. Sup.*, X, 1927, p. 66 : author's italics). Stephenson's ' unanalysed abilities,' due to ' non-fractional factors,' are of the same type (*Brit. J. Psych.*, XXX, 1937, pp. 100–102).

entire columns can be calculated from these, and entered at the same time. This method of setting out such tables was suggested in the working instructions for my own procedure (e.g., Table VIII, *Factors of the Mind*, p. 478); but it can also be adopted to illustrate the method proposed by Holzinger (cf., for example, Thomson, 11, table on p. 339).

1. *The Addition Method.* As in applying the principle of simple summation to bipolar analysis, so here the addition method requires us, after adding each column, to divide their totals by an appropriate divisor. The quotients will then give us the saturation coefficients for the first factor. With group factor analysis, however, we have to omit the diagonal submatrices, since these depend on group factors as well as on the general factor. This omission will naturally entail a corresponding modification in the mode of calculating the divisors. Instead of one divisor for all the column totals (as in bipolar analysis) we shall have to use as many different divisors as there are omitted submatrices—in the present example, three.

Let us use the letters A, B, and C to denote the subtotals of the three non-diagonal submatrices, as indicated in the table. Then, as may be easily shown (see Appendix, p. 73), for the first two columns (i.e., for the coefficients entering into the submatrices A and B) the formula for the divisor will be $\sqrt{C}\left\{\sqrt{\frac{A}{B}}+\sqrt{\frac{B}{A}}\right\}=\sqrt{C}\left\{\frac{A+B}{\sqrt{AB}}\right\}$. It will be observed that the figures required for the terms inside the braces are the sums of those submatrices whose column totals are to be divided. Thus, in the case of the two columns comprising submatrices A and B, the divisor will be $\sqrt{\cdot 48}\left\{\sqrt{\frac{1\cdot 59}{\cdot 81}}+\sqrt{\frac{\cdot 81}{1\cdot 59}}\right\}=1\cdot 4652$. Similarly, for the next two columns, it will be $\sqrt{\cdot 81}\left\{\sqrt{\frac{1\cdot 59}{\cdot 48}}+\sqrt{\frac{\cdot 48}{1\cdot 59}}\right\}=2\cdot 1325$; and for the last two, $\sqrt{1\cdot 59}\left\{\sqrt{\frac{\cdot 81}{\cdot 48}}+\sqrt{\frac{\cdot 48}{\cdot 81}}\right\}=2\cdot 6087$. On dividing each column total by the appropriate divisor as just computed, we obtain the first row of saturations shown in the last line but one.

The theoretical matrix of rank one, to be fitted to the matrix of observed correlations, is obtained in the usual way by multiplying each saturation coefficient by all the rest. This is analogous to the device first proposed by Pearson for fitting a theoretical matrix of rank one to an observed ' contingency table ' in cases of ' manifold association ' (cf. Yule (1910), *Introduction to Statistics*, p. 64, eq. 1, and refs.). To test goodness of fit we may accordingly adopt the chi-squared procedure, employed by Pearson for such cases. The residuals are found by subtracting each theoretical correlation from the corresponding observed correlation; then squared, summed, and multiplied by the number of persons tested. This was the method adopted in earlier investigations both with bipolar and with group factor analysis. However, in the latter case we need to locate the pairs of tests which appear to indicate an overlap of factors. Hence for most purposes it is simpler and more effective to test each residual separately by the standard error of the observed correlation. If there is no overlap, the residuals in the three side submatrices should *all* be non-significant.[1]

Checks.—(i) With this procedure, as a little algebra will show,[2] each column of residuals

[1] The only significance test that could be rigorously defended on statistical grounds would require the theoretical values to be fitted by the method of least squares. We could then derive a likelihood ratio from the products of the factor variances or latent roots. But the requisite formula would, I fancy, be too complex for practical use. Mr. J. R. Robertson has, at my suggestion, applied the simpler significance-tests to artificial tables constructed by using Tippett's random numbers, and finds those described in the text to be reasonably satisfactory, provided the observed correlations are not large and the number of persons tested (N) is not small. When the observed correlations are large, I have recommended one or other of two methods of correction. In earlier investigations the residuals were augmented by using Yule's formula for partial correlation. In later investigations, carried out after 1925, we have commonly used Fisher's z-transformation, and multiplied the square-sum by $(N-3)$ instead of N. (This method of testing discrepancies between correlations has also been used for more general purposes, e.g., by Tippett himself, *Methods of Statistics*, p. 144.)

[2] Proofs of all these checks are given in *Laboratory Notes* cited above.

(excluding, of course, the residuals in the diagonal submatrices) should add up to zero. We can thus apply the same kind of check that we use in other forms of summational analysis. By way of illustration I have calculated, in Table II below, the residuals for the first column.

TABLE II. ADDITION METHOD. CHECKING THE RESIDUALS

Tests	Correlations		Residuals
	Observed	Calculated	
1 and 3	·40	·5511	− ·1511
4	·26	·3764	− ·1164
5	·44	·2857	+ ·1543
6	·30	·1868	+ ·1132
Sum	1·40	1·4000	·0000

(ii) The second check is also analogous to that used in simple summation. (*a*) If we add all the saturations, except those for a particular block, their sum should be equal to the divisor for that block. Thus the sum of the first four saturations is 1·6380 + ·9707 = 2·6087, which is identical with the last divisor (see Table I). (*b*) This check may be put in a slightly different form. If, instead of adding the subtotals for the saturations, we multiply them, their product should be equal to the total of the submatrix common to the two blocks involved. Thus 1·6380 × ·9707 = 1·59 = *A*. Similarly with the other sets of saturations.

2. *The Multiplication Method.* With only three factors the multiplication method reduces to a procedure [1] which, at first sight, seems even simpler than the addition method.

Instead of adding the row- (or column-) sums of the separate submatrices, we multiply them : e.g., for test 1 (which enters into submatrices *A* and *B*), instead of taking the sum (·66 + ·74) = 1·40, and then dividing it by a divisor calculated as above, we take the product of those figures (·66 × ·74) = ·4884, and divide this product by the sum of the remaining submatrix (*C*), namely, ·48. The saturation coefficient is the square root of this quotient. The calculation of the several saturations thus proceeds as follows. Test 1, $\sqrt{(·66 \times ·74)} \div ·48 = 1·0087$; test 2, $\sqrt{(·93 \times ·07)} \div ·48 = ·3682$; test 3, $\sqrt{(·95 \times ·28)} \div ·81 = ·5731$; etc.

We note (i) that the saturation for the first test here rises to over 1·0000, an impossible figure, and (ii) that the saturation for the second test falls to an unduly low figure, ·3682. Generally speaking, it will be seen that, wherever owing to errors of sampling the observed coefficients happen to be too large (or too small) for a precise hierarchical fit, the process of multiplication tends to enlarge (or to diminish) the resulting saturation to an exaggerated degree.

Neither of the checks described above is applicable here. Thus the residuals for the first test are as follows :

TABLE III. MULTIPLICATION METHOD. CHECKING THE RESIDUALS

Tests	Correlations		Residuals
	Observed	Calculated	
1 and 2	·40	·578	− ·178
3	·26	·401	− ·141
4	·44	·302	+ ·138
5	·30	·197	+ ·103
Sum	1·40	1·478	− ·078

[1] Holzinger, I think, does not explicitly illustrate how his own method is to be used in the case of three group factors only. The steps required, however, can be readily deduced from his procedure (described below) for the case of more than three factors; but with three factors only, the actual working becomes greatly simplified. See Holzinger's *Student Manual*, p. 26.

It will also be noted that the saturations given by my own procedure yield a much closer fit. The mean discrepancy (i.e., the average residual) is ·092 with the addition method, and ·107 with the multiplication method.

B. MORE THAN THREE GROUP FACTORS

1. *The Addition Method.* The principle described in section A1 above can readily be extended to the case in which more than three group factors are required.[1] But, as in factorizing individual correlations, so in factorizing the sums obtained by condensation : whenever there are more than three given figures to be factorized, i.e., when we are concerned with a square matrix of an order larger than 3 × 3, we must insert estimated values for the missing cells in the leading diagonal.

As an illustrative example let us take the following table. To avoid introducing large decimal fractions, which would not only take up undue space but also be harder to follow, I have chosen artificial figures ; and the values for the ' observed ' correlations are made to differ very slightly from those which would have been obtained by multiplying the saturation coefficients assumed when constructing the table.

As before, we first rearrange, partition, and condense the matrix of individual correlations (Table IV). The row- and column-sums, and the totals for each column, are then calculated, as for Table I.

TABLE IV. CALCULATION OF SATURATIONS FOR THE BASIC FACTOR

Matrix of Observed Correlations Partitioned and Condensed

Test	1	2	(*a*) Sum	3	4	(*b*) Sum	5	6	(*c*) Sum	7	8	(*d*) Sum
1	—	—	—	·63	·55	1·18	·44	·36	·80	·27	·18	·45
2	—	—	—	·57	·48	1·05	·40	·32	·72	·23	·16	·39
(*a*) Sum ...	—	—	—	1·20	1·03	2·23	·84	·68	1·52	·50	·34	·84
3	·63	·57	1·20	—	—	—	·35	·27	·62	·21	·14	·35
4	·55	·48	1·03	—	—	—	·30	·24	·54	·18	·11	·29
(*b*) Sum ...	1·18	1·05	2·23	—	—	—	·65	·51	1·16	·39	·25	·64
5	·44	·40	·84	·35	·30	·65	—	—	—	·16	·10	·26
6	·36	·32	·68	·27	·24	·51	—	—	—	·12	·09	·21
(*c*) Sum ...	·80	·72	1·52	·62	·54	1·16	—	—	—	·28	·19	·47
7	·27	·23	·50	·21	·18	·39	·16	·12	·28	—	—	—
8	·18	·16	·34	·14	·11	·25	·10	·09	·19	—	—	—
(*d*) Sum ...	·45	·39	·84	·35	·29	·64	·26	·21	·47	—	—	—
Grand Total	2·43	2·16	4·59	2·17	1·86	4·03	1·75	1·40	3·15	1·17	·78	1·95
Divisors ...	2·70			3·10			3·50			3·90		
Saturations	·90	·80	1·70	·70	·60	1·30	·50	·40	·90	·30	·20	·50

We have now to factorize the sums of the submatrices. This involves factorizing a condensed 4 × 4 matrix instead of the original 8 × 8 matrix. The sums, already calculated

[1] In describing my procedure Thomson observes that " Burt only gives his formula for three groups " (11, p. 341). However, in my paper on ' Factor Analysis by Submatrices,' the use of the method with more than three groups was briefly described and illustrated (8, pp. 350f., section headed " *More than three group factors*," and equation iv, p. 352 ; cf. also 9, p. 482, note in small print). It had previously been employed in one or two L.C.C. investigations and in several unpublished theses by research students.

in the body of Table IV, are for convenience set out afresh in Table V. To factorize Table V by simple summation, we must first complete the table by inserting suitable figures in the diagonal. These, of course, should be the squares of the saturations we are about to calculate. But, as we do not yet know the saturations, we must proceed by successive approximation ; in this way we shall ultimately reach the figures shown in the diagonal within brackets. The calculation of the saturations is then accomplished in the usual manner, by dividing each column total by the square root of the grand total. (It will be noted that the figures inserted in brackets agree with the squares of the saturations so found : $2{\cdot}89 = 1{\cdot}70^2$, etc.)

TABLE V. ADDITION METHOD

Calculation of Divisors from Condensed Correlation Matrix

Group	Sum of Submatrices (*a*)	(*b*)	(*c*)	(*d*)	Total
(*a*)	(2·89)	2·23	1·52	·84	7·48
(*b*)	2·23	(1·69)	1·16	·64	5·72
(*c*)	1·52	1·16	(·81)	·47	3·96
(*d*)	·84	·64	·47	(·25)	2·20
Divisor 4·4	7·48	5·72	3·96	2·20	19·36 = 4·4
Sats. for Sums ...	1·70	1·30	·90	·50	4·40
Reference letter ...	S_a	S_b	S_c	S_d	T
Divisors	2·70	3·10	3·50	3·90	13·20

Let us call the ' saturations for sums ' S_a, S_b, S_c, and S_d respectively, and their total T. Then the divisor for the columns on which S_a is based will be $T - S_a$; for those on which S_b is based, $T - S_b$; and so on. The total of the divisors $= (m - 1)T = 3 \times 4{\cdot}40 = 13{\cdot}20$, where m as before denotes the number of groups (here four). On dividing the totals for the test-columns, as given in the initial table (Table IV), we obtain the eight saturation coefficients shown in the bottom line : thus $2{\cdot}43 \div 2{\cdot}70 = {\cdot}90$, etc.

Checks. The same two checks can be used as before : viz., (1) the sum of all the saturations except those belonging to a given block should be identical with the divisor for that block ; (2) the residuals should add up to zero.

2. *The Multiplication Method.* As we have seen, the bifactor method depends essentially on the multiplication of sums rather than on the addition of sums. The working procedure may be summarized as follows.[1]

First, the totals of the sums of the submatrices are found as in Table V, but *without* inserting estimated figures for the diagonal cells : thus, $2{\cdot}23 + 1{\cdot}52 + {\cdot}84 = 4{\cdot}59$. These totals are as a matter of fact given at the foot of the columns headed (*a*), (*b*), (*c*), and (*d*) in Table IV. But to make the working quite clear, the sums have been copied out and totalled afresh in Table VI. To find the divisors, the totals so obtained are subtracted from *half* the grand total ($6{\cdot}86 - 4{\cdot}59 = 2{\cdot}27$, etc.).

TABLE VI. MULTIPLICATION METHOD

Calculation of Divisors for Bifactor Analysis

Group	Sum of Submatrices (*a*)	(*b*)	(*c*)	(*d*)	Total
(*a*)	—	2·23	1·52	·84	4·59
(*b*)	2·23	—	1·16	·64	4·03
(*c*)	1·52	1·16	—	·47	3·15
(*d*)	·84	·64	·47	—	1·95
Total	4·59	4·03	3·15	1·95	$13{\cdot}72 \div 2 = 6{\cdot}86$
Divisors	2·27	2·83	3·71	4·91	13·72

[1] See 10, Appendix C, pp. 328–341.

A second supplementary table (Table VII) is now constructed from Table IV as follows. The first figure in each row of Table VII consists of the total of all the sums in the corresponding row of Table IV, except the first sum (·80 + ·45 = 1·25) ; the next figure of the total of all the sums except the first and second ; and so on if there are more totals left to calculate. In the present case the second ' total ' is simply the last figure in the corresponding row of Table IV, i.e., ·45. It is convenient to enter each of these ' diminishing sums ' under the letter corresponding to the omitted sum.

TABLE VII.
TABLE OF DIMINISHING SUMS

Group	(*a*)	(*b*)	(*c*)
Test 1	—	1·25	·45
2	—	1·11	·39
3	·97	—	·35
4	·83	—	·29
5	·91	·26	—
6	·72	·21	—
7	·67	·28	—
8	·44	·19	—

TABLE VIII.
PRODUCTS FOR NUMERATOR

Group	(*a*)	(*b*)	(*c*)	Sum
Test 1	—	1·4750	·3600	1·8350
2	—	1·1655	·2808	1·4463
3	·1640	—	·2170	1·3810
4	·8549	—	·1566	1·0115
5	·7644	·1690	—	·9334
6	·4896	·1071	—	·5967
7	·3350	·1092	—	·4442
8	·1496	·0475	—	·1971

These ' diminishing sums ' have now to be multiplied by the figures that were successively omitted, i.e., the first and second subtotals from Table IV (the last subtotal, for *d*, will not be required, because there is no entry corresponding to it in Table VII [1]). We obtain the products shown in Table VIII. The products in each row are then summed ; and the sums divided by the group divisors already found in Table VI. The square root of the quotient gives the saturation for the test (see Table IX). It will be remarked that, although the ' observed ' correlations differ only very slightly from the exact hypothetical values, as constructed from the saturation assumed, the bifactor method does not reproduce those saturations so accurately as the group factor method (' addition method '). Further, the checks are no longer applicable.

TABLE IX. CALCULATION OF SATURATIONS

Test	Sum of Products (Numerator)	Divisor (Denominator)	Quotients	Saturations (Square Root)	True Values
1	1·8350	2·27	·8084	·8991	·90
2	1·4463	2·27	·6371	·7982	·80
3	1·3810	2·83	·4880	·6986	·70
4	1·0115	2·83	·3574	·5978	·60
5	·9334	3·71	·2516	·5015	·50
6	·5967	3·71	·1609	·4011	·40
7	·4442	4·91	·0905	·3008	·30
8	·1971	4·91	·0401	·2003	·20

Comparison of Alternative Methods. The theoretical grounds for preferring what I have called the addition method will, I hope, now be clear. With the addition method we are in effect basing our estimate of the saturation coefficient on a process of averaging. In fact, the expression for the saturation coefficient, with this mode of calculation, might be regarded as a kind of a weighted average, the divisor being, as usual, the sum of the weights. This is the natural principle to adopt in smoothing

[1] The whole process is equivalent to multiplying each sum by the total of all the rest, and then halving the sum of the products. The ' total of all the rest ' can be obtained by deducting in turn each sum from the column total. Thus for the first row we should have the following figures :

	(*a*)	(*b*)	(*c*)	(*d*)	Sum
Sums ...	— ...	1·25 ...	1·63 ...	1·48	—
Products ...	— ...	1·4750 ...	1·3040 ...	0·8910 ...	3·6700 = 2 × 1·8350

out the effects of the sampling fluctuations that must affect each observed coefficient. The multiplication formula, on the other hand, is a direct algebraic deduction from a postulated equation ; and the deduction is strictly applicable only when that postulated equation is exactly fulfilled. With an empirical table this can never be the case. Consequently, the method is exposed to all the risks incurred when errors are implicitly multiplied instead of being added.[1]

A few words are perhaps necessary to indicate how the procedures so far described differ from that adopted by Spearman and his followers. There are three points of divergence which deserve some comment. (i) The principal difference arises from the fact that Spearman always distrusted any attempt to assess the general factor on the basis of internal evidence. He therefore preferred to use ' objective tests ' of ' pure *g*,' and to eliminate *g* by the ordinary formula for partial correlation. This would seem to be a perfectly legitimate procedure, provided (*a*) that the general factor we wish to eliminate is identifiable with *g* as independently assessed (i.e., with what is ordinarily described as ' general intelligence '), and (*b*) that we possess a satisfactory set of tests for assessing ' pure *g*.' Otherwise it would seem wiser to take the information supplied from the entire battery of tests into account. (ii) But further, Spearman himself thought it necessary, on methodological grounds, to deal with only one group factor at a time. Thus, in re-examining the evidence for the various group factors adduced by other investigators, instead of analysing an entire correlation table which might need to be partitioned for three or more group factors, he preferred to confine himself to a table, or part of a table, which could be reduced to a 2×2 form, consisting of inter-correlations and cross-correlations between two groups or tests only. And here, instead of looking for a group factor in each of the two groups, he insisted, as before, on treating one of the groups as consisting of ' reference values,' dependent solely on *g*, while the second group was examined for the presence of a possible group factor. (iii) Thirdly, Spearman objected to my attempt to derive group factors by summing an entire submatrix of ' like ' tests for the same reasons that he objected to my attempt to derive a general factor by summing the entire correlation table. In both cases factors were treated as averages ; and in his view this was a mistake (6, pp. 61f.). Strictly speaking, his own formula was applicable only to the case in which there were *two* ' reference abilities ' and in which *two* ' abilities ' were to be examined for specific correlation (6, p. 223, and Appendix, p. xxi). Consequently, " some complication," as he says, " is introduced when the available abilities are more than two in number, for not more than two can be used at the same time. . . . The more thorough way is to try out every possible pair." [2]

His method of looking for a single group factor only I shall discuss more fully in a moment. Meanwhile, let us consider how the group factor method can be adapted to the case where there are only two group factors.

C. TWO GROUP FACTORS

Frequency of Paired Group Factors. The need to deal with two group factors occurs in three main types of investigation.

(i) It occurs in dealing with those correlation tables that are reducible to a 2×2 form, with which, as we have just seen, Spearman preferred to work. The procedure that he proposed to substitute is by no means satisfactory : few of his critics accepted his demand that, in every given analysis, all group factors but one should be dispensed with. Indeed, more often than not, the tests which he employed to eliminate *g* seemed themselves to involve a second group factor of their own. For example, in dealing with Miss Rogers' study of mathematical ability, he reduced the tests

[1] There are other differences between the two methods : e.g., Holzinger's regression equation for estimating factor-measurements (*Preliminary Reports*, No. 8, p. 7) includes only those tests that contain the group factor ; mine uses all the tests for all the factors. I have discussed this and other differences more fully in (8). As stated above, full proofs of the relevant formulæ are given in the roneo'd *Notes* already cited. But once again, I imagine, the reader will have no difficulty in working them out for himself. In my notes, instead of the notation of the Appendix, I suggested writing $a_1, a_2, \ldots b_1, b_2, \ldots$ etc., for the hypothetical saturations in groups *a*, *b*, etc. The expressions for the working equations and their proofs can then be readily written down.

[2] In practice, however, he considered it both legitimate and sufficient to " pool together all those abilities that enter into the same set." But in principle there seems very little difference between the proposal to " pool " and the proposal to calculate factors by " summation " or " averaging." I should add, however, that the formula which Spearman suggested for use when pooling was not altogether correct, even if his own assumptions are accepted (see this *Journal*, I, pp. 96f.).

employed to two main groups : (*a*) ' arithmetical tests,' to be examined for a possible group factor, and (*b*) ' reference values,' to serve as tests of pure *g*. Now the latter included Analogies, Opposites, Completion, and Thorndike's Test of Reading. All of them were verbal tests ; and it seemed highly likely that they contained a verbal factor as well as ' pure *g*.' The same was true of his discussion of Collar's research—a somewhat similar study of the arithmetical factor based on work in L.C.C. schools : here again it seemed clear that each of the two contrasted groups of tests included its own specific group factor.[1]

(ii) In the analysis of physical traits a similar situation often occurred. In the earliest empirical table which seemed to call for an analysis in terms of group factors—namely, the Pearson-Macdonell table for head and body measurements to which I have already referred—the requisite division was clearly into two main clusters of ' like traits.' The procedure which I very tentatively proposed drew sharp criticisms both from Spearman and from Pearson himself. In later work with group factors, therefore, to avoid objections directed primarily against the technique adopted rather than against the conclusions drawn, we generally tried to arrange the battery of tests or traits so that there should, if possible, always be at least three group factors. In psychological work, it was quite easy to secure this condition : indeed, in early work with mental tests, where the number of persons tested was small, and the analysis could not pretend to isolate more than a couple of factors in each test, this threefold grouping was the type most commonly encountered. On the other hand, in the field of physical measurement, the study of ' body types '—particularly of the recurring contrast between the leptomorphic (or asthenic) type and the eurymorphic (or pyknic) type—constantly called for analysis in terms of two group factors.

(iii) Later, when larger populations became available for mental testing, it was often found that many of the group factors hitherto discovered were apt to get subdivided in their turn ; and the division was usually by pairs. Indeed, an extremely common type of factor-pattern proved to be that in which the whole set of traits involves two broad group factors, and then each of these splits into two subfactors, and so on. Hence it seemed urgently necessary to devise an acceptable method which should at once be convenient in practice and free from serious theoretical objections.

Now, even those investigators who have attempted to deal with several group factors in the same correlation table appear generally to assume that, unless a battery includes at least one test which involves no group factors at all but only the general factor, it is impossible to determine the factor-saturations with two groups only : just as " three variables at least are always necessary for a Spearman ' two factor ' analysis," so (it seems to be supposed) three groups at least are always necessary for a group factor analysis. Thus, in illustrating how a set of seven tests can be factorized to give one general and two group factors, Holzinger constructs his artificial correlation table so that the seventh test has zero saturations for every factor except the first.[2] Actually, however, there are several ways in which the difficulty can be overcome. Here I shall content myself with what in practice has proved to be the simplest and most convenient.

Example. In contributions previously published I have always applied the method to empirical figures obtained from actual researches. But for purposes of illustration, the errors and discrepancies inevitably involved in dealing with observed coefficients tend to obscure the underlying principles. Here, therefore, I shall take an artificial table of correlations, constructed to give simple and exact calculations. A convenient set of figures, obtained by rounding off observed results from tests of

[1] Rogers, A. (1923), ' Tests of Mathematical Ability,' Columbia Univ. Cont., No. 130, Table IX, pp. 56–57 ; Collar, D. J. (1920), ' A Statistical Survey of Arithmetical Ability,' *Brit. J. Psych.*, XI, pp. 135f. Miss Rogers' data seemed to me to afford admirable support for my contention. Actually it included a *third* group of tests, mainly perceptual, which Spearman omitted. It was thus possible to show that the verbal factor, which appeared with my ' two-group-factor analysis ' of the small table of correlations selected by Spearman, was identical with the verbal factor which appeared with a ' three-group-factor analysis,' when the *whole* of Miss Rogers' table was employed. With two group factors the chief objection urged by Spearman and others was that the factors and their saturations could not be uniquely determined. But with three group factors, of course, there could be no doubt about the uniqueness of the figures obtained. Spearman, however, further maintained that, in any case, no such thing as a group factor for verbal ability existed (cf. *Abilities of Man*, p. 237). In later editions of his book he altered his view, and added this factor to his list. The reader who desires to try out the methods described in the remainder of this article will find in Miss Rogers' figures an empirical table admirably adapted for the purpose.

[2] Holzinger, H. *Student Manual*, p. 26.

educational abilities,[1] is shown in Table X. The coefficients, as we shall see in a moment, have been built up by using one basic and two non-overlapping group factors ; the table therefore has a rank of three. Accordingly, by successive approximation, a computer, given nothing but the test-correlations, should be able to determine exactly the appropriate values for the self-correlations (shown in Table X within brackets).

TABLE X. OBSERVED CORRELATIONS

Test	1	2	3	4	5	6	7
1. Verb. Int.	(·82)	·76	·48	·29	·63	·45	·36
2. Sci. 1 (Gen.)	·76	(·80)	·52	·32	·56	·40	·32
3. Sci. 2 (Mech.)	·48	·52	(·34)	·21	·35	·25	·20
4. Dict.	·29	·32	·21	(·13)	·21	·15	·12
5. Non-Verb. Int.	·63	·56	·35	·21	(·53)	·43	·32
6. Mech. Ass.	·45	·40	·25	·15	·43	(·41)	·28
7. Sci. 3 (Pract.)	·36	·32	·20	·12	·32	·28	(·20)

TABLE XI. FACTOR-SATURATIONS

Tests	A. Bipolar Analysis				B. Group Factor Analysis			
	I	II	III	Square-Sum	*G*	*V*	*P*	Square-Sum
1. Verb. Int. ...	·8824	+·0625	+·1939	·8200	·90	·10	—	·82
2. Sci. 1 (Gen.) ...	·8567	+·2500	−·0596	·8000	·80	·40	—	·80
3. Sci. 2 (Mech.) ...	·5471	+·1874	−·0747	·3400	·50	·30	—	·34
4. Dict.	·3329	+·1249	−·0596	·1300	·30	·20	—	·13
Sum (Verbal) ...	2·6191	+·6248	·0000	—	2·50	1·00	—	—
5. Non-Verb. Int....	·7054	−·1562	+·0894	·5300	·70	—	·20	·53
6. Mech. Ass. ...	·5518	−·3124	−·0894	·4100	·50	—	·40	·41
7. Sci. 3 (Pract.) ...	·4190	—·1562	·0000	·2000	·40	—	·20	·20
Sum (Practical) ...	1·6762	−·6248	·0000	—	1·60	—	·80	—
Square-Sum ...	2·9003	·2635	·662	3·2300	2·69	·30	·24	3·23
Contrib. to Variance (per cent.)	41·4	3·8	·9	46·1	38·4	4·3	3·4	46·1

As usual we begin with a preliminary ' bipolar analysis.' The saturations for the general and two bipolar factors, found in this way, are shown in Table XI A. The figures indicate that a clear line of division can be drawn between the first four tests and the last three. As with the Pearson-Macdonell table, the cross-correlations between the four traits forming the first group and the three traits forming the second group fit a hierarchical arrangement. These cross-correlations are shown as a separate table in Table XII. Since they form a matrix of rank one, they can be regarded as due to a single common factor only, namely, the ' basic ' factor.

[1] The tests are described below on p. 62.

(i) *Determining the Saturations for the Basic Factor.* We now proceed to analyse this rectangular submatrix in much the same way as we should analyse a complete square matrix of observed correlations, when extracting a single factor only. Since, however, the submatrix is asymmetrical, the sums of the columns will no longer be identical with the sums of the rows. We must therefore consider how we can find a suitable divisor. To take the square root of the grand total, as we do with a symmetrical table, would, as a general rule, be quite indefensible. For one thing, there are four tests furnishing column-sums, but only three furnishing row-sums ; and this fact of itself must make the column-sums smaller. Secondly, quite apart from the difference in the number of tests in either group, we have no right to assume that the factor-saturations in each group will be equally balanced : as we shall find in a moment, the average saturation for the first three tests is only 0·53, and for the last four 0·63.

To allow for these two peculiarities we must therefore use two divisors instead of one. For a first approximation the simplest procedure is to select the divisors so that in the group factor analysis the sum of the factor-saturations for the first three tests will have the same proportion to the sum of the factor-saturations for the last four tests as they do in the bipolar analysis.

Now let d_1 and d_2 be the two divisors required for calculating the group factor-saturations from the row- and column-sums ; t_1 and t_2 the sums of the group factor-saturations for the corresponding groups ; t the sum of all the correlations in the submatrix ; and s_1 and s_2 the sums of the factor-saturations obtained for the first and second groups of tests by the preliminary bipolar analysis. Then

$$t_1 = \frac{t}{d_1}, \text{ and } t_2 = \frac{t}{d_2}.$$

But $t = t_1 t_2 = d_1 d_2$. Hence $d_1 = t_2$, and $d_2 = t_1$.

Now, since $\frac{d_2}{d_1} = \frac{t_1}{t_2} = \frac{s_1}{s_2}$, we have $d_2 = \frac{s_1}{s_2} d_1$; and $t = d_1 d_2 = \frac{s_1}{s_2} d_1^2$.

Hence $$d_1 = \sqrt{\left\{\frac{s_2}{s_1} t\right\}} \text{ and } d_2 = \sqrt{\left\{\frac{s_1}{s_2} t\right\}}.$$

Thus all we have to do is to multiply the total (t) by the proportion of the two sums as given by the bipolar analysis (s_1/s_2 or s_2/s_1), and take the square root. Accordingly, inserting the appropriate figures from Tables XI A and XII, we find

$$d_1 = \sqrt{\left\{\frac{2{\cdot}6191}{1{\cdot}6762} \times 4{\cdot}00\right\}} = 2{\cdot}50,$$

$$\text{and } d_2 = \sqrt{\left\{\frac{1{\cdot}6762}{2{\cdot}6191} \times 4{\cdot}00\right\}} = 1{\cdot}60.$$

On dividing the row-sums by 2·50 and the column-sums by 1·60, we obtain the saturations shown on the extreme right and at the foot of Table XII. We can check the figures along the same lines as before by noting that the sum of the saturations for tests 1 to 3, viz., $t_1 = 1{\cdot}60 = d_2$; and the sum for tests 4 to 7, namely, $t_2 = 2{\cdot}50 = d_1$. Further, $1{\cdot}60 \times 2{\cdot}50 = 4{\cdot}00$.

TABLE XII. CALCULATION OF SATURATIONS FOR BASIC FACTOR (G)

Non-Verbal Tests	Verbal Tests 1	2	3	4	Row-Sums	Sats.
5	·63	·56	·35	·21	1·75	·70
6	·45	·40	·25	·15	1·25	·50
7	·36	·32	·20	·12	1·00	·40
Column-Sums	1·44	1·28	·80	·48	4·00	1·60
Sats.	·90	·80	·50	·30	2·50	—

Divisors : for Column-Sums, 1·60 ; for Row-Sums, 2·50.

TABLE XIII. CALCULATION OF SATURATIONS FOR GROUP FACTORS (*P*)

Test	5	6	7	Sums (for checking)
Saturations (Basic Factor) ...	·70	·50	·40	—
Test 5	·53	·43	·32	1·28
Sat. ·70	·49	·35	·28	1·12
Residuals	·04	·08	·04	·16
Test 6	·43	·41	·28	1·12
Sat. ·50	·35	·25	·20	·80
Residuals	·08	·16	·08	·32
Test 7	·32	·28	·20	·80
Sat. ·40	·28	·20	·16	·64
Residuals	·04	·08	·04	·16
Totals of Residuals	·16	·32	·16	$·64 = ·80^2$
Saturations (Group Factor) ...	·20	·40	·20	·80

Divisor for Totals $= \sqrt{·64} = ·80$.

(ii) *Determining the Saturations for the Supplementary Group Factors.* The next step is to eliminate the effects of this general or basic factor from the observed correlations. We may begin by reconstructing a theoretical hierarchy for the smallest group from the saturations just obtained for tests 5 to 7, and deduct these theoretical values from the inter-correlations actually observed, as shown in Table XIII. The first set of residuals for these three tests are then factorized by simple summation in the usual way. We can then do the same with the inter-correlations for the larger group, viz., that containing tests 1 to 4. The complete set of factor-saturations thus obtained is shown in Table XI B.

Now our hypothesis is that there are only two factors entering into the inter-correlations for each group of tests. If that hypothesis is correct, then, within each group, the *first* set of residuals (those remaining after the basic factor has been eliminated) should form a submatrix of rank one only (with the smaller group a 3×3 submatrix, with the larger a 4×4 submatrix). Hence, with an ideal table of correlations like the present the real test for the suitability of our basic factor-saturations is that the *second* set of residuals (those remaining after the appropriate group factor has been eliminated) should vanish identically. With the figures given in Table X this actually happens. In the more general case the fit obtained by using the basic factor given by the first approximation would not be so perfect ; further approximations may therefore be required. The routine procedure would consist in subtracting the effects of the group factors alone from the observed inter-correlations for both groups ; and then substituting the residuals for the observed inter-correlations in the original table. The entire 7×7 table can then be regarded as an approximate hierarchy, and can therefore be factorized by simple summation to obtain improved values for the basic factor. However, this routine procedure may be considerably abridged by making suitable adjustments as they suggest themselves by simple inspection. Alternatively, with submatrices of this size we can write out an algebraic expression for the sum of the residuals in terms of the factor saturations treated as unknowns, set this sum equal to zero, and solve for the unknowns. It is unnecessary to illustrate this procedure here, since it will seldom be needed with any actual table.

With an empirical table we cannot expect to find the second set of residuals summing to *exactly* zero. In such a case, therefore, the best procedure would be one that minimized the sum of the squares of the residuals. Again, it is not difficult to construct algebraic equations which will yield a working formula for obtaining the required results by direct calculation. But the calculation would be unnecessarily laborious. For most practical purposes it will be simplest to proceed by successive approximation in the way just suggested, modifying the factor-saturations obtained with

the first trial in order to diminish both sets of second residuals so far as possible. Even so, in the vast majority of cases, unless extremely large populations have been tested, the sampling errors will be big enough to render any elaborate adjustments of this kind superfluous.

For a simple example of this procedure applied to an actual correlation table, the reader may refer to the analysis of the Pearson-Macdonell table quoted in the preceding issue of this *Journal* (12, p. 118, Table IX ; cf. also Table XV B below). But, as a rule, the matrix of non-overlapping group factors merely serves as an intermediate step towards obtaining the final matrix of overlapping group factors ; and consequently in most published articles only the latter is printed.

D. ONE GROUP FACTOR

For the sake of completeness I shall conclude this section by considering how the foregoing principles are to be applied to a correlation table in which there appears to be only a single group factor. Spearman, as we have seen, always doubted the legitimacy of trying to extract more than one group factor from the same table at the same time. When confronted with a table in which the possibility of several group factors called for consideration (as in researches planned, not by himself, but by other investigators), he still preferred to take one group factor at a time, even if this meant discarding a large portion of the table. His method was to select correlations so that these should, so far as possible, relate to two groups of tests only. One group was to be regarded as providing " reference values for *g* " : these, it was assumed, would involve no further factors. The other group would include, or at any rate *might* include, in addition to the general factor, *g*, the supplementary group factor, whose existence he proposed to examine (see above p. 53).

Spearman's procedure, however, sought to investigate simply the presence or absence of this group factor (6, ch. XIII). Even when he admitted its presence, he never went on to estimate its saturation coefficients. Indeed, he devised no formula, and suggested no method, for such determinations.[1]

[1] The procedure which Spearman eventually substituted for group factor analysis is worth describing in some detail. In the current text-books on factorial methods, his criterion is always described, but no explanation is given to show how he actually used it (cf. 10, pp. 68, 112f. ; 11, pp. 12f. ; Thurstone, *Multiple Factor Analysis*, ch. XII). His own account (6, pp. 223 and xxi) is by no means easy to follow. The fullest explanation of the method is to be found in Hargreaves' *Monograph* on ' The Faculty of Imagination ' (*loc. cit. sup.*, 1927, pp. 18–24), where he describes in detail an attempt to isolate a single group factor of ' fluency ' from a correlation table including 8 tests of ' intelligence ' and 6 tests of ' imagination,' to which tests of Memory, Speed, and Perseveration were added later.

As the investigation was carried out in L.C.C. schools, a preliminary conference was held, as usual, between the supervisor (Spearman), the L.C.C. psychologist (myself), and the head teachers concerned ; and the correspondence-file furnishes an instructive account of the alternative suggestions. In discussing the statistical procedure, I had recommended " an extended group factor method " (virtually that described in Section II B above), on the grounds that, not only my own students, but also " Miss Carey, Bradford, Collar, and others had tried it with some success," and that it " dealt most directly with the problem, as envisaged by teachers." Spearman, however, replied that " the true criterion (the tetrad difference criterion) has now been provided " : " even the older [inter-columnar] criterion has shown that the group factors, formerly claimed, have no real existence ; and this is now confirmed by the true criterion." " To estimate the size of the supposed group factors merely by accidental departures from a calculated hierarchy is," he maintained, " very unsafe. . . . Every supposed group factor must be scrutinized by itself."

In his *Monograph* Hargreaves accordingly discusses each of the conceivable factors or ' faculties ' one by one. He begins by applying the tetrad difference criterion to the tests of Intelligence and Imagination separately ; and so shows, or claims to show, that " each of the two groups of tests—Intelligence and Imagination (' Fluency ')—contains a general factor and no second general or group factors." Then, applying the same criterion to the two groups " taken together," he states that " a group factor or second general factor is found." Thus, *one* of the two general factors previously referred to " must be *compound*." Rather summarily, he dismisses the possibility that the Intelligence tests might contain a second factor, " because of evidence that is accumulating against the existence of a group factor in intelligence tests " (he does not specify this evidence, except that on a later page he briefly refers to the possibility of a ' linguistic ' group factor in such tests, similar to the ' linguistic ability ' I had reported, and rejects it as " most improbable "). Hence in his view the only conclusion remaining is that the initial correlation table contains a *single* group

TABLE XIV. CORRELATIONS BETWEEN PHYSICAL MEASUREMENTS

One Group Factor : Calculation of Basic Factor

Trait	1	2	3	4	Sum	5	6	7	Sum
1. Foot... ...	(·821)	·736	·759	·797	3·113	·363	·339	·206	·908
2. Height ...	·736	(·731)	·661	·800	2·928	·345	·340	·183	·868
3. Finger ...	·759	·661	(·714)	·846	2·980	·321	·301	·150	·772
4. Forearm ...	·797	·800	·846	(·803)	3·246	·289	·305	·135	·729
Sum ...	3·113	2·928	2·980	3·246	12·267	1·318	1·285	·674	3·277
Saturation ...	—	—	—	—	—	·376	·367	·193	·936
5. Face B. ...	·363	·345	·321	·289	1·318	—	·395	·618	—
6. Head L. ...	·339	·340	·301	·305	1·285	·395	—	·402	—
7. Head B. ...	·206	·183	·150	·135	·674	·618	·402	—	—
Sum ...	·908	·868	·772	·729	3·277	—	—	—	—
Total ...	4·021	3·796	3·752	3·975	15·544	—	—	—	—
Saturation ...	·906	·855	·845	·896	3·502	—	—	—	—

Divisor for traits 1–4 = $\frac{15{\cdot}544}{\sqrt{12{\cdot}267}} = 4{\cdot}4381$ Divisor for traits 5–7 = 3·502.

To illustrate the procedure I propose, we may take the Pearson-Macdonell 7 × 7 table, which was analysed in the previous issue of this *Journal* in terms of two group factors superposed upon a single basic factor (see Table XIV). We may imagine a follower of Spearman enquiring whether we cannot after all dispense with one of these two group factors. Can we not assume that there is just a single basic factor determining the general growth of the body and its limbs, and a special group factor confined to the measurements of head and face? Let us therefore see what kind of a fit this hypothesis yields, by calculating the requisite figures according to the principle of simple summation.[2]

We begin by determining the basic saturations for the four body-traits (as we may call them). The first step is to find the sums of the columns of inter- and cross-correlations for these four traits. As usual, the final sums must include the self-correlations[3] : for these, trial values are first inserted, and then corrected in the ordinary way by substituting the squares of the self-correlations so found. The divisor for each column of seven figures is found by dividing the total of all four columns (15·544) by the square root of the sum of the inter-correlations, including the self-correlations (i.e., by $\sqrt{12{\cdot}267} = 3{\cdot}502$). We thus obtain for the saturations 4·021 ÷ 4·4381 = ·906, etc. (see last line of Table XIV).

factor of ' fluency,' entering into the tests of Imagination in addition to the all-pervasive *g*. Since these six tests were shown to obey the tetrad difference criterion, it is inferred that the ' compound general factor ' must be what was later called a ' bifid factor ' or ' bifactor ' : i.e., as he says, the two factors must enter into the several tests " in exactly the same proportion." To clinch the conclusion so reached, and determine more precisely the essential nature of the so-called fluency factor, he employs Yule's formula for partial correlation (a method used by Miss Spielman and myself in studying ' creative imagination,' cf. *A Study in Vocational Guidance*, 1926, pp. 23f.).

In the light of his own experience Hargreaves himself eventually admits that he is not altogether content with this procedure. Where single group factors are to be separately disproved, he thinks it satisfactory enough. But " when several factors are likely to be operative, the [tetrad difference] criterion seems to have a limited value : it informs us of the existence of more than one factor, but cannot help to separate the factors out." As to the application of Yule's formula, " apart from the validity of such a proceeding, the mathematical labour involved, when more than one factor has to be eliminated, and when several tests are involved, becomes very great."

[2] See Appendix, I, p. 73 below, for the derivation of the formula employed.

[3] Spearman's procedure involves the assumption that the self-correlations are all equal to 1·000 : see this *Journal*, I, p. 96. When the cross-correlations include instances of specific overlap, the saturations for the basic factor should in theory be based on the inter-correlations only.

To find the basic factor saturations for the last three traits, we first sum the three 4-figure columns of cross-correlations. For these the divisor is the sum of the four saturations just found (3·502). We thus obtain 1·318 ÷ 3·502 = ·376, etc. The usual checks can be applied : for example, the residuals of each of the four 7-figure columns and of each of the three 4-figure columns should evidently add to zero.

The saturations for the single group factor are obtained by subtracting the theoretical values for the inter-correlations between the three head traits (calculated by multiplying the three saturations just found) from the observed inter-correlations. Since for these traits there are only three observed inter-correlations, and therefore only three residuals, an exact fit can here be obtained. The saturations are computed by the familiar formula $r_{58} = \sqrt{\frac{r_{56}\, r_{57}}{r_{67}}}$, etc., where r_{56}, r_{57}, r_{67} denote the three residuals : e.g., for trait 5 we have $= \frac{\cdot 257 \times \cdot 545}{\cdot 331} = \sqrt{\cdot 4229} = \cdot 650$, etc.

The factor-saturations so found are shown in Table XV. For comparison I give the figures obtained on the assumption that there are *two* group factors (from 12, p. 118, Table IX). The question now arises : which of the two alternative assumptions furnishes the better fit ?

Judged by the absolute amount of discrepancy, the hypothesis of two group factors certainly yields the closest fit. The numerical total of the residuals comes to 0·257 with two group factors, and 0·502—nearly twice as much—with one group factor. Judged according to the principle of least squares, the difference is still more marked ; the square-sum in the former case is 0·0045, in the latter 0·0218.

TABLE XV. FACTOR-SATURATIONS

Trait	A. One Group Factor		B. Two Group Factors		
	Size	Head	Size	Body	Head
1. Foot...	·906	—	·719	·501	—
2. Height	·854	—	·688	·464	—
3. Finger	·845	—	·612	·582	—
4. Forearm	·896	—	·578	·816	—
5. Face B.	·376	·650	·508	—	·505
6. Head L.	·367	·395	·495	—	·284
7. Head B.	·193	·839	·260	—	·962
Square-Sum	3·379	1·282	2·270	1·471	1·261
Contribution to Variance (per cent.)	48·3	18·3	32·4	21·0	18·0

But are these differences significant ? As the observed correlations were calculated by Pearson's tetrachoric procedure, it is not easy to assign exact figures for the standard errors of the residuals. Macdonell's estimate puts them in the neighbourhood of ± 0·015. Roughly speaking, therefore, any residual that is numerically larger than 0·030 may be regarded as significant. With two group factors no residual reaches this figure. With one group factor seven out of 21 residuals exceed it. The standard deviation with the two-group factor hypothesis is 0·0147 (just about what we should expect by chance) ; with the one-group factor hypothesis it is more than twice as large, namely, 0·0312.

In point of fact, a preliminary inspection of the table of observed correlations might have led us to anticipate some such result. I have arranged the traits so that the rectangular submatrix of cross-correlations (i.e., the correlations of traits 1–4 with traits 5–7) should conform, so far as possible, to a hierarchical order. It will be noted that they do so with remarkable closeness. If traits 1–4 were also due to the same general factor and nothing more, the inter-correlations for these traits should also be hierarchical, and (what is more important) should exhibit the same order. But this is not the case. The trait which has the *largest* sum for its *inter*-correlations (Forearm) has the *smallest* sum for its *cross*-correlations.

If our critic suggests that perhaps after all we might have been wiser to take, not the three head measurements, but the four body measurements as containing the sole additional group factor, then his suggestion would be still more strongly contradicted by the same argument : for, as simple inspection will show, Head Breadth has the *largest* sum for its inter-correlations, and the *smallest* sum for its cross-correlations. Thus, *both* sets of inter-correlations demand a group factor of their own.

In practice I am inclined to believe that, with correlation tables which are expected to reveal only a single group factor, the investigator will, as a rule, be better advised to start by allowing for two group factors, and then prove that the saturations for one are non-significant. In general, however, I should regard an experiment which yields a correlation table showing one group factor only (or indeed no more than two group factors) as badly planned. Unless the number of persons tested or measured is exceptionally large, and the standard errors in consequence exceptionally small (as they are here), it is usually possible to fit such tables with so many different group factor patterns that the results are bound to be inconclusive.[1]

III. OVERLAPPING GROUP FACTORS

Rotation. When the number of persons tested is large, and especially when the assortment of tests has been chosen to fulfil a practical need rather than to solve a theoretical problem, a pattern of non-overlapping factors is likely to give a somewhat incomplete fit to the correlations observed : in such investigations, as experience shows, we nearly always discover residuals indicative of overlap [2] in the outlying submatrices. But in any case we cannot really assume that an empirically derived factor matrix must always contain a number of saturations that are *exactly* zero. Hence it is desirable to have ready for use some simple arithmetical procedure which will enable us to readjust the schematic pattern of non-overlapping factors so as to yield a closer overall fit.

For this purpose the most obvious device is to rotate the bipolar factor matrix obtained by simple or weighted summation to fit the factor-pattern obtained by the group factor analysis.[3] The use of orthogonal transformations is, of course, a familiar procedure in mathematics ; and the idea of arithmetically rotating factors in this way was first suggested by Garnett (5). Graphical methods of rotation have been freely used by Thurstone ; but, as Reyburn and others have pointed out, the substitution of graphical for numerical methods is not only less exact, but appears to allow wide room for subjective influences (13).

Example. To illustrate the need for this supplementary procedure and a convenient method of carrying out the calculations let us consider the artificial table of correlations set out on p. 63. It may be regarded as based on a set of 10 cognitive tests applied to a sample of 500 cases (Table XVI).

In the original investigation, planned as part of an experiment in educational and vocational guidance,[4] a series of written, oral, and practical tests was applied to several batches of boys, mostly

[1] It was for this reason that I have not illustrated my arguments in this section by an example drawn from mental measurement. Earlier discussions (e.g., 6, pp. 224–241) provide numerous instances in which one group factor only is discussed. But in nearly every case the probable error is too large to make any clear comparisons profitable. Further, in many cases the wide differences in reliability tend of themselves to introduce a misleading parallelism into the coefficients. For instance, in Hargreaves' table (cited above) the reliability of the tests employed varies from 0·23 to 0·92 ; and this of itself tends to produce an appearance of consistent proportionality between the inter-correlations and the cross-correlations.

[2] The existence of ' overlap ' between group factors was first noted in the analysis of emotional traits, but proved to be a regular feature in analysing correlations between educational abilities (4, pp. 59–60). Spearman uses the term in a somewhat different sense : where I had described two or more tests as involving a *group* factor, he preferred to say that they showed an overlap between their respective *specific* factors (6, p. 82).

[3] Cf. 8, p. 358 ; 9, p. 308. A more fully worked example is given in *Brit. J. Educ. Psych.*, IX, 1939, pp. 55–65.

[4] The investigation (amongst others) was described in a Report to the Chief Inspector of the L.C.C. Education Department on *Examinations for Junior County Scholarships and Trade Schools*, and I am indebted to the Council for permission to make use of the results.

between the ages of 12½ and 14½, and attending trade schools in or near London. The written tests consisted of two tests of Intelligence, verbal and non-verbal, two papers on ' Arithmetical Problems ' and ' Practical Geometry ' respectively, and finally two papers on ' Science ' requiring brief written answers mostly in essay form—the first, a ' general ' paper, dealing mainly with elementary physics and including interpretations of diagrams, the second, dealing more specifically with ' Mechanics ' and involving some appreciation of formulæ. These written tests were set to 543 boys, usually as part of a terminal examination at the institute they were attending. Two further tests, Dictation and Mental Arithmetic, were given orally, but required written replies : they were given to the same boys, with the exception of two smaller schools. Two practical tests were included in the scheme—an Assembly test and a test of Practical Science (physics, mechanics, and the like) ; but these could only be given to a limited number of pupils, amounting in all to 457.

The correlations were calculated for each group separately ; and the average correlations (weighted for numbers and reliability) were factorized by the method to be described in a moment. To obtain a more convenient table for purposes of illustration, I have here rounded off the saturation coefficients to a single significant figure ; and have then reconstructed an artificial table of ' observed ' correlations (Table XVI), which, as a matter of fact, closely resembles the table obtained from the largest group. By taking an example based on an exact set of saturations, we shall have a definite check for the figures ultimately reached by the procedure proposed.

Bipolar Analysis. We begin with a preliminary bipolar analysis. Four factors have been extracted, three of which we may regard as statistically significant. The saturations so obtained are shown in Table XVII A. The interpretation of the three significant factors is quite obvious. The first is a factor of ' general ability '—a weighted average of all the tests ; the second divides the tests into non-numerical (tests 1–7) and numerical (tests 8–10) ; the third divides the non-numerical tests into verbal (tests 1–4) and non-verbal, i.e., practical (tests 5–7). And the factor pattern as a whole suggests that the correlations could probably be fitted by means of a comprehensive basic factor supplemented by three smaller group factors. However, in the lateral submatrices of Table XVI several of the coefficients are much too large for a perfect hierarchical arrangement. Hence we may anticipate some degree of overlapping between the several group factors.

Analysis into Non-Overlapping Group Factors. As before, we may take the bipolar analysis as indicating how the correlation table should be rearranged and where the lines of division between the several groups should fall. The rearrangement has already been carried out in printing the correlations in Table XVI ; and the table itself has been partitioned accordingly. The next step is to factorize it in terms of three non-overlapping group factors in order to discover the general frame-work.

An experienced computer would probably realize at once that a satisfactory fit could only be obtained by successive approximation, and would probably start by reducing the more conspicuous enlargements due to the overlapping at the outset. Alternatively he might seek to base the group factor analysis solely on those parts of the lateral submatrices that conform fairly well to hierarchical requirements. The beginner will prefer a simpler and more automatic procedure ; and here therefore I shall assume that the method described above (p. 48) is to be applied to the observed correlations taken just as they stand.

The results of factorizing the table by the method described in Section II A are shown in Table XVII B. This second factor-pattern gets rid of negative saturations ; but manifestly will not fit the table of observed correlations as well as the bipolar matrix. It is easy to see, for example, without actually reconstructing a theoretical table of correlations from the factor-saturations, that the residuals for tests 2 and 3 with 10 and for tests 2 and 9 with 6 will be fully significant. What we now require therefore is a readjusted group factor matrix derived from the previous bipolar matrix (Table XVII A), but conforming to the schematic pattern exhibited by Table XVII B.

We could, of course, use these two tables to calculate a transformation matrix : if this transformation matrix proved to be strictly orthogonal, then the fit obtained by the rotated factors would be as good as that obtained direct from Table XVII A itself, and the new factors would themselves be orthogonal. However, the correlation tables reconstructed from the group and bipolar factors (sections A and B) would be by no means identical ; and consequently the transformation matrix would be far from orthogonal. We might go on to modify the transformation

TABLE XVI. OBSERVED CORRELATIONS

Test	1	2	3	4	Sum	5	6	7	Sum	8	9	10	Sum
1. Verb. Int. ...	(·82)	·76	·48	·29	—	·63	·44	·37	1·44	·72	·54	·21	1·47
2. Sci. 1 (Gen.) ...	·76	(·84)	·52	·32	—	·60	·44	·40	1·44	·64	·52	·28	1·44
3. Sci. 2 (Mech.)...	·48	·52	(·35)	·21	—	·35	·22	·23	·80	·44	·35	·25	1·04
4. Dict.	·29	·32	·21	(·13)	—	·21	·13	·14	·48	·24	·18	·12	·54
Sum	2·35	2·44	1·56	·95	7·30	1·79	1·23	1·14	4·16	2·04	1·59	·86	4·49
5. Non-Verb. Int.	·63	·60	·35	·21	1·79	(·53)	·43	·32	—	·56	·46	·14	1·16
6. Mech. Ass. ...	·44	·44	·22	·13	1·23	·43	(·42)	·27	—	·40	·38	·07	·85
7. Sci. 3 (Pract.) ...	·37	·40	·23	·14	1·14	·32	·27	(·21)	—	·32	·28	·11	·71
Sum	1·44	1·44	·80	·48	4·16	1·28	1·12	·80	3·20	1·28	1·12	·32	2·72
8. Prob. Arith. ...	·72	·64	·44	·24	2·04	·56	·40	·32	1·28	(·80)	·68	·40	—
9. Prac. Geom. ...	·54	·52	·35	·18	1·59	·46	·38	·28	1·12	·68	(·65)	·42	—
10. Arith. Oral ...	·21	·28	·25	·12	·86	·14	·07	·11	·32	·40	·42	(·49)	—
Sum	1·47	1·44	1·04	·54	4·49	1·16	·85	·71	2·72	—	—	—	—
Col. Total	2·91	2·88	1·84	1·02	8·65	2·95	2·08	1·85	6·88	3·32	2·71	1·18	7·21
Divisors			3·301					4·333				4·208	
Saturations ...	·882	·872	·557	·309	2·620	·681	·480	·427	1·588	·789	·644	·280	1·713

matrix, so as to make it orthogonal. But this procedure would, as a rule, be rather laborious ; and in any case is suited only to more experienced computers.[1]

Calculation of Rotation Matrix from Condensed Factor Matrices. If, however, instead of taking the initial table of observed correlations, we proceed to reconstruct a new correlation table from Table XVII B as it stands, and if we then subject this reconstructed table to a bipolar analysis, we should have two factor matrices, bipolar and group, which furnished *exactly* the same set of correlations. The rotation matrix for transforming the first factor matrix into the second would then be perfectly orthogonal. Since, however, we are concerned with factor matrices containing 4 columns only, the order of the rotation matrix will be 4×4 ; and consequently, to calculate such a rotation matrix, we do not need group- and bipolar-matrices containing 10 rows ; 4 will be sufficient. Let us therefore condense the group factor matrix into a 4×4 matrix ; from this let us reconstruct a symmetrical product-sum matrix, by the procedure regularly used for reconstructing a theoretical correlation table. The 4×4 matrix so obtained will then, as a rule, be identical with the table we should get if we condensed the reconstructed 10×10 correlation table to a 4×4 table of summed correlations.

Accordingly, with the present investigation we can begin by splitting the first group of 4 tests into two subgroups, the first containing test 1 only and the second tests 2, 3, and 4. With the other three groups we can take the sums as they stand. We shall then work with the sums of the factor-saturations for these four groups (see Tables XVII B and XVIII A). The 4×4 matrix of summed correlations constructed from these sums is shown in Table XVIII B. This is now subjected to a bipolar analysis ; and the saturations so obtained are given in Table XVIII C. These should be identical with the sums of the bipolar factor-saturations we should have extracted from the complete 10×10 correlation matrix.[2]

It is now an easy matter to calculate the transformation matrix, required for rotating the figures in Table XVIII A to those in Table XVIII C. There is no need to determine an inverse. The fact that the sums of the non-overlapping group factors include a number of zeros renders a direct calculation, by the ordinary method of solving linear equations, perfectly simple (for the method of calculation see Appendix, p. 75).

The transformation matrix thus obtained is given in Table XIX A. The reader can easily satisfy himself that it is orthogonal.

Rotation of Original Factors. (*a*) *First Approximation.* We now use this transformation matrix to postmultiply the *original* bipolar-matrix (Table XVII A). This yields the set of overlapping group factors as shown in Table XX A. The pattern of saturations is fairly satisfactory ; and, if the number tested had been small, it could be accepted as it stands.

It reveals several clear instances of positive overlap, e.g., tests 7 and 10 in the second factor (*V*), test 9 in the third factor (*P*). But it also yields three negative saturations over 0·100 (for tests 3, 8, and 10). Such negative values are apt to occur in sections where there is an appreciable positive

[1] It has, as a matter of fact, been used in several research-theses (e.g., Miss Stevenson's analysis of tests applied in the Navy during the war). The orthogonalizing process is best carried out as follows. The general principle will be to restrict the largest adjustments to those rows or figures of the transformation which will multiply the smallest columns in the bipolar matrix. To begin with, we should leave unchanged the first row in the transformation matrix which multiplies the first column of the bipolar matrix ; we should alter the least important figure in the second row, so that the product-sum of the first and second rows is zero ; we should then modify the two least important figures in the third row so that its product-sum with the first two rows is zero ; and so on. Finally, for the last and least important row of all we substitute figures which will make the sums of the squares in each column equal to unity.

[2] As in the ' sum method ' of factor analysis (this *Journal*, II, pp. 62f.). Certain cautions have to be exercised in regard to the reversal of signs. The beginner is advised in his first efforts to factorize the complete 10×10 artificial correlation table as a check : with certain modes of pooling and with the ordinary rules for sign reversal, he may find that the totals for later factors differ slightly when obtained from the full table and from the summed table respectively. As a rule, such minor divergences will not appreciably affect the final results ; and in any case, those results will generally be treated as a first approximation. Hence the second approximation can be trusted to eliminate any defects due to the provisional mode of pooling.

TABLE XVII. FACTOR-SATURATIONS FOR BIPOLAR AND NON-OVERLAPPING GROUP FACTORS

Test	A. Bipolar Factors I	II	III	IV	Square-Sum	B. Group Factors *G*	*V*	*P*	*N*	Square-Sum
1. Verb. Int.	·8513	·1968	·0085	·2378	·8200	·8816	·0579	—	—	·7806
Sum	·8513	·1968	·0085	·2378	—	·8816	·0579	—	—	—
2. Sci. 1 (Gen.)	·8610	·2442	·1737	—·0945	·8400	·8725	·2381	—	—	·8180
3. Sci. 2 (Mech.) ...	·5503	·0405	·2134	·0075	·3500	·5574	·1734	—	—	·3480
4. Dict.	·3188	·0839	·1460	·0036	·1300	·3090	·2184	—	—	·1432
Sum	1·7301	·3686	·5331	—·0834	—	1·7389	·6299	—	—	—
5. Non-Verb. Int. ...	·6846	·1796	—·1612	·0558	·5300	·6807	—	·2420	—	·5220
6. Mech. Ass.	·5179	·1625	—·3309	—·1262	·4200	·4800	—	·4344	—	·4191
7. Sci. 3 (Pract.) ...	·4289	·1287	—·0495	—·0840	·2100	·4267	—	·1480	—	·2040
Sum	1·6314	·4708	—·5416	—·1544	—	1·5875	—	·8244	—	—
8. Prob. Arith.	·8416	—·2174	—·0738	·1978	·8000	·7890	—	—	·3712	·7603
9. Prac. Geom.	·7218	—·3192	—·1589	—·0433	·6500	·6440	—	—	·4615	·6277
10. Arith. Oral	·4030	—·4996	·2328	—·1545	·4900	·2804	—	—	·5901	·4268
Sum	1·9664	—1·0362	·0000	·0000	—	1·7134	—	—	1·4228	—
Square-Sum	4·1827	·5810	·3196	·1569	5·2400	3·9364	·1379	·2692	·6990	5·0425
Contribution to Variance (per cent.)	41·8	5·8	3·2	1·6	52·4	39·4	1·4	2·7	7·0	50·4

overlap in the final saturations, and where the provisional non-overlapping matrix gives zero saturations : the negative saturations are then arithmetically necessary to balance the positive saturation that represents the overlap. By taking the correlations in the lateral submatrices just as they stand, we have in effect treated the augmented coefficients as non-significant results of sampling fluctuation. That means that our saturations represent averages, and the residuals represent non-significant deviations, positive and negative, about such averages. At this point, therefore, it is desirable to make a rough test of the significance of the larger positive and negative saturations which take the place of the previous zero saturations, and, if they approach the size of the probable error, proceed to a further approximation.

(*b*) *Second Approximation.* There are various ways in which an improved rotation can be effected. Two may be described as most convenient for general purposes.

(i) The rotated matrix just obtained reveals where the overlaps occur and what is their probable size. We can therefore go back to the original correlation matrix, and deduct from the augmented coefficients the amount apparently due to the overlap. The analysis into non-overlapping group factors can then be repeated with the smoothed correlation table ; and the whole of the subsequent calculations carried out afresh. This is probably the simplest routine for the beginner.[1]

(ii) An alternative method is to modify two or three figures in the rotation matrix, so as to suppress the larger negative saturations, and then orthogonalize the columns afresh. Thus the negative saturations in the second and the third factors are obviously due to the large negative figures in the second and third columns of the rotation matrix (— ·8379 and — ·8598) : these can very readily be reduced. Minor adjustments must then be made in other figures to ensure that the product sums of the different pairs of columns are exactly zero ; and each of the readjusted columns must then be normalized to ensure that the sum of its squares is 1·000.

Final Results. An adjusted rotation matrix obtained in this way [2] is shown in Table XIX B. This is used, as before, to postmultiply the original bipolar-matrix (Table XVII A). The resulting factor saturations are shown in Table XX B. For comparison I append in the last four columns the factor-saturations from which the original correlation matrix was actually constructed. It will be observed that the agreement is remarkably close. With three minor exceptions the discrepancies between the calculated values and the true values are all less than $\pm$ 0·015 ; and, of course, with a further series of approximations we could get as close as we desire.

The interpretation of the factors is fairly obvious : the first is a ' basic ' factor of ' general ability ' ; the rest represent ' special abilities ' for verbal, practical, and numerical work. The instances of overlap are perfectly intelligible, when we recall the nature of the tests.

Constancy of Group Factor Matrices. In any field of enquiry where the investigators may wish to assign a concrete meaning to each of the factors extracted, it is desirable to determine the factors in such a way that their interpretation shall remain the same, even when they are calculated afresh from a modified battery, to which certain tests have been added, or from which certain tests have been omitted. In previous accounts of the group factor method, I have claimed that it will fulfil this

[1] As noted above, the more experienced computer would probably have observed at the very outset which of the coefficients disturbing the hierarchical arrangement showed significant augmentations, and would have made tentative deductions before undertaking the first group factor analysis. Even with a brief inspection of Table XVI, it was fairly obvious that test 10, for example, shows exaggerated correlations with tests 2, 3, and 4. When these coefficients have been reduced, several smaller augmentations are readily detected (e.g., tests 2 and 8 with tests 5, 6, and 7).

[2] This and the remaining calculations were made for me by a former research student, Mr. F. R. Llewellyn, who knew nothing of the true values of the saturation coefficients or of how the original correlation table had been constructed. I felt that, were I myself to make the requisite adjustments, I might be consciously or unconsciously influenced by my previous knowledge of the true values —or at least that it might be supposed that some such influence had operated. Having reached the figures shown in Table XVII B, Mr. Llewellyn began to suspect that he was working on an artificial problem. In the hope of suppressing the negative saturation for test 6, he spontaneously carried out a further approximation. Working with six decimal figures he then obtained the true values to within five decimal places (including the negative saturation).

TABLE XVIII. FACTOR ANALYSIS OF PRODUCT-SUMS

Tests	A. Summed Group Factor-Saturations (G_s)				B. Product-Sums ($R = G_sG_s'$)				C. Bipolar Analysis of R (B_s)			
	G	*V*	*P*	*N*	1	2–4	5–7	8–10	I	II	III	IV
1	·8816	·0579	·0000	·0000	3·1644	1·6358	2·7846	3·0054	·8506	·1568	·1062	·1453
2–4	1·7389	·6299	·0000	·0000	1·6358	·9042	1·3755	1·4846	1·7352	·4177	·4628	—·1453
5–7	1·5875	·0000	·8244	·0000	2·7846	1·3755	3·1996	2·7200	1·6300	·4680	—·5690	·0000
8–10	1·7134	·0000	·0000	1·4228	3·0054	1·4846	2·7200	4·9601	1·9681	—1·0425	·0000	·0000

TABLE XIX. ROTATION MATRICES

Factors	A. First Approximation ($T_s = G_s^{-1} B_s$)				B. Second Approximation			
	G	*V*	*P*	*N*	*G*	*V*	*P*	*N*
I	·9575	·1113	·1334	·2303	·9296	·2071	·1769	·2482
II	·1641	·2100	·2517	—·9304	·1917	·1105	·2013	—·9542
III	·0882	·4914	—·8598	—·1063	—·0778	·8415	—·5337	—·0308
IV	·2199	—·8379	—·4236	—·2648	·3051	—·4865	—·8020	—·1645

TABLE XX. FACTOR-SATURATIONS FOR OVERLAPPING GROUP FACTORS

Test	A. First Approximation					B. Second Approximation					C. True Values				
	G	*V*	*P*	*N*	Square-Sum	*G*	*V*	*P*	*N*	Square-Sum	*G*	*V*	*P*	*N*	Square-Sum
1. Verb. Int. ...	·901	−·056	·053	−·049	·820	·901	·086	−·010	−·014	·820	·900	·100	—	—	·820
2. Sci. 1 (Gen.) ...	·859	·312	·068	−·019	·840	·805	·398	·187	−·009	·840	·800	·400	·200	—	·840
3. Sci. 2 (Mech.)	·554	·169	−·103	·067	·350	·505	·294	−·014	·090	·350	·500	·300	—	·100	·350
4. Dict.	·331	·122	−·063	−·020	·130	·301	·195	−·008	−·059	·130	·300	·200	—	—	·130
5. Non-Verb. Int.	·683	−·013	·251	−·008	·530	·700	−·001	·198	−·006	·530	·700	—	·200	—	·530
6. Mech. Ass. ...	·466	·033	·448	·035	·420	·499	−·095	·402	·004	·420	·500	−·100	·400	—	·420
7. Sci. 3 (Pract.) ...	·409	·120	·168	·007	·210	·402	·102	·195	−·001	·210	·400	·100	·200	—	·210
8. Prob. Arith. ...	·807	−·153	·036	·353	·800	·807	·009	−·014	·386	·800	·800	—	—	·400	·800
9. Prac. Geom. ...	·614	−·029	·171	·492	·650	·609	−·008	·183	·496	·650	·600	—	·200	·500	·650
10. Arith. Oral ...	·289	·185	−·205	·576	·490	·212	·300	−·030	·595	·490	·200	·300	—	·600	·490
Square-Sum ...	3·930	·216	·388	·706	5·240	3·773	·398	·309	·760	5·240	3·730	·410	·328	·780	5·240
Contrib. to Variance (per cent.)	39·3	2·2	3·9	7·1	52·4	37·7	4·0	3·1	7·6	52·4	37·3	4·1	3·3	7·8	52·4

requirement within the limits of the probable error. May I therefore conclude my description by a brief illustration of this point?

Let us take the last correlation table (Table XVI), and omit the final group of tests, i.e., the 'numerical' tests, 8, 9, and 10. The remaining 7 × 7 table can now be factorized by the method described in Section C for the case of two group factors only. The saturations obtained first by preliminary bipolar analysis and then the final analysis in terms of non-overlapping group factors are set out in Table XXI B.

TABLE XXI. FACTOR ANALYSIS OF CURTAILED BATTERY

Tests	A. Bipolar Factors			B. Group Factors		
	I	II	III	*G*	*V*	*P*
1. Verb. Int.	·874	·076	·225	·881	·121	—
2. Sci. 1 (Gen.) ...	·895	·138	—·144	·881	·219	—
3. Sci. 2 (Mech.) ...	·542	·211	—·042	·489	·345	—
4. Dict.	·330	·140	—·039	·294	·223	—
5. Non-Verb. Int. ...	·708	—·139	·098	·704	—	·178
6. Mech. Ass.	·542	—·354	—·029	·484	—	·451
7. Sci. 3 (Pract.) ...	·447	—·072	—·069	·448	—	·093

On comparing the figures reached with those given in Table XX B, it is obvious that the agreement is exceedingly close. In only one instance (test 2) does the discrepancy exceed 0·100. We could, if we wished, attempt a further approximation; or, if the small residuals were statistically significant, seek a slightly closer fit by a rotation which would substitute small saturations for the zeros.

The initial correlation table, as we know, includes several instances of overlapping. It is this overlapping that has caused the small discrepancies. With only two group factors we can scarcely expect to detect or to allow for it with any degree of accuracy. The ideal procedure would be, as before, to invoke the principle of least squares. The calculations so entailed would be quite practicable, but exceedingly lengthy; and in general the results would differ but little from those achieved by the simpler method described above.

With artificial tables, containing no overlapping whatever, the saturations reached by the group factor analysis would necessarily be identical throughout, no matter how many groups of tests we added or removed (provided, of course, at least two groups remained), and no matter what tests we dropped from the several groups (provided, of course, at least two tests were left in each). This will be obvious if the reader turns back to Section II C, and considers the analysis of Table X, which in point of fact was constructed from the group factor saturations shown in Table XX C with the overlapping eliminated (cf. Table XI B). In such cases therefore the figures would remain perfectly constant or invariant.

Simple Structure. Thurstone's conception of a unique 'simple structure' is in many ways closely similar to what some of us in earlier publications had described as a 'group factor pattern.' A 'simple structure' should, we are told, satisfy three conditions: if V denotes the factor matrix of r columns, then "(1) each row of V should have at least one zero [saturation]; (2) each column of V should have at least r zeros; (3) for every pair of columns there should be at least r traits whose entries vanish in one column but not the other."[1] It will be observed that the second of these conditions precludes any 'general' or 'basic' factor, since by definition such a factor must have a non-zero saturation for every trait in its column.

In addition there are bound to be further differences in the distribution and saturations of the other factors, resulting partly from the differences in the conditions and partly from differences in method. The 'primary factors' making up a 'simple structure' are derived from a preliminary 'centroid' analysis by a series of 'rotations' (not necessarily orthogonal). The 'rotations' are carried out, not arithmetically, but by a graphical procedure which is usually lengthy and inevitably somewhat rough; nor is any use made of the sign-pattern of the centroid matrix to determine the essential limits of the several group factors. Their limits are in effect largely dependent on the choice of the computer at each stage, as he selects this or that pair or triplet of variables for the next rotation.

[1] *Vectors of Mind* (1935), p. 156.

Both the similarities and the differences can readily be observed in the concrete examples where both methods of analysis have been applied to the same correlation table, e.g., where Professor Thurstone has re-analysed a set of traits, previously factorized by the group factor method, and where a set of traits previously factorized in terms of Thurstone's simple structure has been re-analysed by the group factor method.[1] Wherever a strict orthogonal rotation would inevitably yield a general factor, Professor Thurstone prefers to analyse the correlation matrix in terms of an ' oblique ' simple structure, i.e., in terms of correlated factors ; and then (as he has suggested in his more recent contributions) a further analysis can be undertaken, if we wish, to extract a set of ' second order ' factors, one of which will be the general factor.

Thus with Thurstone's procedure three consecutive analyses are required to do what can frequently be done by a single analysis, if we seek a simple group factor pattern.[2] In more complex cases, where there is considerable overlapping, there seems no reason to insist on conditions 1 and 3. Thus, instead of assigning the zero (or non-significant) saturations solely to satisfy certain *a priori* requirements, expressed arbitrarily for all cases in terms of the empirical number of factors (a number which is largely an accidental result of the size of the sample), my own method is to decide their location by the aid of the preliminary bipolar pattern. That pattern, of course, is in turn really determined by the explicit or implicit hypothesis, which has guided the investigator in choosing his particular sets of tests or traits at the very outset of his experiment (provided that his hypothesis is applicable to the data so procured).

There are other less obvious differences depending partly on our mathematical and partly on our psychological preconceptions. But these would require discussion at some length, and must therefore be deferred to a later paper. Here I have said enough to show that there are clear differences both in method and in results. May I add that I do not wish to deny that there may be problems for which Thurstone's procedure is the best, just as there may be other problems for which the bifactor method may be more appropriate ?

In conclusion let me repeat that this article has been primarily concerned with practical working methods only. There are a number of theoretical issues (e.g., the precise relation between the bipolar classification and the group factor classification, which is by no means simple in the case of the smaller bipolar factors, the criteria for choosing between a bipolar and a group factor hypothesis, the method of planning experiments for group factor analysis, etc.) which I have no space to touch on here.

Advantages of the Group Factor Method. In my previous discussions of various factorial procedures I have suggested that the factors extracted should, if possible, fulfil three main conditions : paucity, unequivocality, and stability. From our survey of the group factor method, it may, I think, be fairly claimed that its results will in general satisfy these requirements. First, the values given for the group factor-saturations will be (within the limits imposed by the sampling error) either positive or zero ; hence each of the traits or tests to be factorized will be accounted for in terms of the minimum number of uncorrelated factors. Secondly, each factor will be (within the limits of sampling) uniquely determined. Thirdly, the factors will remain stable, if not absolutely invariant, even when the battery of tests or traits is modified, e.g., when a comparatively small battery is enlarged by the addition of more tests or more groups of tests, or when a large battery is curtailed by the omission of tests. Occasionally, it is true, the results obtained may fail to preserve all three conditions. But such cases are, I believe, exceptional. They occur, in my experience, only where the investigator has selected his battery unwisely, and where, as a consequence, *no* factorial procedure could really be expected to yield satisfactory results.

[1] In *Psychometrika*, XI (1946), p. 20, Professor Thurstone presents an alternative analysis for a table of physical measurements. In *Brit. J. Educ. Psych.*, IX, 1939, pp. 270–276, Dr. Eysenck has carried out a group factor analysis of the correlation table given by Professor Thurstone in his work on *Primary Abilities*.

[2] In the second edition of his book (11, p. 297) Professor Thomson gives a ' second order analysis ' of a 7-variable version of Thurstone's box problem, and shows how it leads to a " general pattern containing a ' general factor *plus* an orthogonal simple structure.' " But this is simply the old basic-plus-group-factor pattern, reached by a more circuitous route. Reached by the group factor method the pattern of saturations is very much the same, but the fit is closer (cf. Burt, C., *Brit. J. Educ. Psych.*, table on p. 47).

IV. SUMMARY AND CONCLUSIONS

1. Many tables of correlations, covariances, or product-sums, particularly in psychology, can be expressed in terms of a simple and intelligible factor-pattern, if the observed coefficients are treated as the result of (i) a small number of positive 'group' factors each entering into a few of the variables only, superposed upon (ii) a single 'basic' factor entering into all the variables. The aim of this article has been to make a comprehensive and systematic survey of the working methods that have proved to be most effective for this purpose.

2. By definition the significant saturations for both the 'basic' and the 'group' factors will have positive values only. Two modes of approach are available for securing this result—an addition method and a multiplication method. Reasons are given for preferring the former. With that method the essential principle is to condense the matrix of observed correlations by simple summation. This yields a direct mode of analysis which is both easy and speedy, and in the simplest cases provides an adequate set of factors after a single analysis.

3. Where there is any doubt about the grouping of the tests, it is advisable to begin with a preliminary bipolar analysis. The pattern of signs so obtained supplies an objective indication for the appropriate partitioning of the table of observed correlations.

4. Slightly different working procedures are described for the cases of one, two, three, and more than three group factors. The simplest and commonest case is that of three group factors. The hypothesis of subdivided factors, however, will generally require the determination of paired group factors ; and with larger tables there is no fixed limit to the number of factors that can be determined.

5. Where the sample is small, it is usually sufficient to seek non-overlapping group factors. Where the sample is large, it will generally be necessary to allow for overlapping. This can best be done by an arithmetical rotation, which appears less liable to subjective influences than graphical rotation. A suitable rotation matrix can be quickly computed by condensing the group factor matrix by summation, and then making a bipolar analysis of the product-moment table reconstructed from these sums.

6. Finally, it is argued that the factors so found are uniquely determined, economical in number, and virtually invariant when the battery is enlarged or diminished.

REFERENCES

1. Burt, C. (1909). 'Experimental Tests of General Intelligence.' *Brit. J. Psych.*, III, 94–177.
2. Simpson, B. R. (1912). 'Correlations of Mental Abilities.' *Columbia University Contributions to Education*, No. 53, Teachers' College.
3. Bickersteth, M. E., and Burt, C. (1916). 'An Analysis of Results of Mental and Scholastic Tests.' *Report Conf. Educ. Assoc.*, 30–37.
4. Burt, C. (1917). *The Distribution and Relations of Educational Abilities.* L.C.C. Report No. 1868. London : P. S. King & Son.
5. Garnett, J. C. M. (1919). 'General Ability, Cleverness, and Purpose.' *Brit. J. Psych.*, IX, 345–366.
6. Spearman, C. (1926). *Abilities of Man.* London : Macmillan.
7. Burt, C. (1936). 'The Analysis of Examination Marks,' ap. Hartog, P., Rhodes, E. C., and Burt, C. *The Marks of Examiners.* London : Methuen.
8. Burt, C. (1938). 'Factor Analysis by Submatrices.' *J. Psych.*, VI, 339–375.
9. Burt, C. (1940). *Factors of the Mind.* London : University of London Press.
10. Holzinger, K. J., and Harman, H. H. (1941). *Factor Analysis.* Chicago : University of Chicago Press.
11. Thomson, G. H. (1946). *The Factorial Analysis of Human Ability* (2nd ed.). London : University of London Press.
12. Burt, C. (1949). 'Alternative Methods of Factor Analysis.' *Brit. J. Psych.* (*Stat. Sect.*), II, 98–121.
13. Reyburn, H. A., and Raath, M. J. (1949). 'Simple Structure : A Critical Examination.' *Brit. J. Psych.* (*Stat. Sect.*), II, 125–133.

APPENDIX

I. *Basic Formulæ*

Postulates and Definitions. Let x_i and y_i be the observed measurements of any given individual, i, for any pair of correlated traits, e.g., his marks as obtained from two mental tests. We shall assume that such measurements can always be expressed as linear functions (weighted sums) of a set of uncorrelated factors. Accordingly, let us write

$$x_i = w_{xb}b_i + w_{xc}c_i + \ldots + w_{xs}s_i , \quad . \quad . \quad . \quad . \quad . \quad \text{(i)}$$

$$y_i = w_{yb}b_i + w_{yc}c_i + \ldots + w_{xt}t_i , \quad . \quad . \quad . \quad . \quad . \quad \text{(ii)}$$

where b_i, c_i, s_i, and t_i are i's hypothetical measurements for the several factors and remain constant for the same individual in all the different tests, and where w_{xb}, w_{xc}, w_{yb}, w_{yc}, etc., are the weights or 'factor-loadings' for the several tests, and remain constant for the same test in all the different individuals, i.e., they denote the proportionate amounts of the factors b, c, etc., in the tests x and y.

Evidently the weights may take positive, negative, or zero values. A factor which has positive weights for some of the tests and negative for the remainder will be called a *bipolar* factor. A factor which has positive weights for some of the tests and zero values for the remainder will be called a *group* factor. A factor which has positive weights for every test will be called a *basic* factor if all (or practically all) the weights for the other factors are positive, but smaller: otherwise (e.g., if there are bipolar factors) it will be called a *general* factor.[1] A factor which has zero weights for all except one particular test will be called a *specific* or 'unique' factor: when the tests are not repeated, it will include the error factor.

The measurements will be rendered comparable, and the ensuing formulæ greatly simplified, if we suppose the x, y, b, c, . . . s, and t are all expressed in unitary standard measure: so that $\Sigma x^2 = \Sigma y^2 = \ldots = 1$; and therefore $w^2_{xb} + w^2_{xc} + \ldots + w^2_{xs} = 1$; and similarly for the weights in y.

With these assumptions the following theorems can readily be deduced.

(i) *Saturation Coefficients.* On substituting for x from equation (i), we have—

$$r_{xb} = \frac{\Sigma xb}{\sqrt{\Sigma x^2}\sqrt{\Sigma b^2}} = \Sigma(w_{xb}b + w_{xc}c + \ldots + w_{xs}s)b = w_{xb} , \quad . \quad . \quad \text{(iii)}$$

since $\Sigma cb, \ldots \Sigma sb$, will be zero (because by hypothesis the factors are uncorrelated), and $\Sigma b^2 = 1$. Similarly, $r_{xc} = w_{xc}$; $r_{yb} = w_{yb}$; $r_{yc} = w_{yc}$. Thus, the correlation between a given test and any one of its component factors turns out to be identical with the weight assigned to that factor in the theoretical construction of the test. (Spearman's statement that the weight is proportional to the *square* of the correlation between the test and the factor springs from a confusion of this case with that of a correlation between two tests where the weight of the common factor is assumed to be the same for both tests.)

(ii) *Product-Theorem.* Similarly, on substituting for x and y from equations (i) and (ii), we have

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = w_{xb}w_{yb} + w_{xc}w_{yc} + \ldots = r_{xb}r_{yb} + r_{xc}r_{yc} + \ldots \quad . \quad \text{(iv)}$$

(iii) *Bipolar Analysis.* Now let us suppose we have n tests dependent on the general or basic factor, b, and on one supplementary factor, c (say), which we will first suppose to be common to *all* the n tests. Then, taking the total of the xth column of correlations and the grand total of all the correlations, we have (from (iv))

$$\frac{\Sigma r_{xk}}{\sqrt{\Sigma\Sigma r_{kk'}}} = \frac{r_{xb}\Sigma r_{kb} + r_{xc}\Sigma r_{kc}}{\sqrt{(\Sigma r_{kb})^2 + (\Sigma r_{kc})^2}} = r_{xb} , \quad (k, k' = 1, 2, \ldots x, \ldots n) \quad . \quad . \quad \text{(v)}$$

provided $\Sigma r_{kc} = 0$. Thus, for equation (v) to hold, c must be a bipolar factor, and the sum of its positive and negative saturations must be zero. This treats the saturation coefficient, so calculated, as a kind of average. Equation (v) gives the ordinary formula for factor analysis by 'simple summation.' It can obviously be generalized to include the case where there is more than one supplementary factor, like c.

(iv) *Group Factor Analysis.* The working methods that follow depend on what I call the 'principle of condensation' (see above, p. 45). The factor matrix and the correlation matrix are first partitioned conformably, according to the grouping of the traits. The column-sections of the factor matrix, and the submatrices of the correlation matrix, are then summed. The matrices of sums so obtained I call 'condensed' matrices. It can then be shown that the condensed correlation

[1] There is no *a priori* reason why bipolar and group factors should not appear in the same factor matrix. All that is necessary is that the extraction of each factor should reduce the rank of the matrix by one: and this, as is easily shown, is a very broad requirement. The methods here described can readily be combined.

matrix (i) will have the same rank as the original table, and (ii) can be deduced from the condensed factor matrix, i.e., from the summed saturations. For, let R_{12} be any rectangular submatrix obtained by partitioning the correlation table ; let $R_{12} = F_1F_2'$, where F_1 and F_2 are the corresponding sections of the factor matrix ; and let w_0' denote the summation operator [1, 1, . . . 1]. Then $w_0'R_{12}w_0 = w_0'F_1F_2'w_0$, where $w_0'F_1$ and $w_0'F_2$ are the sums of the column-sections.[1] Further, if W' denotes the semi-condensation operator (whose rows consist of 1's or 0's), then $W_0'RW_0 = W_0'FF'W_0$.

(*a*) *One Group Factor.* Let us now suppose that c is common only to a limited group of similar tests. And for simplicity, let us take the case of four tests only, with the first two tests alone involving the group factor c. Expressing the inter- and cross-correlations in terms of the factor-saturations, the whole correlation table, and the sums which result from its condensation, may be written as in Table XXII. Here the capital letters denote sums of factor-saturations, i.e.,

$$B_1 = \sum_1^2 r_{kb}\,,\quad B_3 = \sum_3^4 r_{kb}\,,\quad C_1 = \sum_1^2 r_{kc}\,.$$

To find the *basic* factor-saturations from the column totals ($r_{1b}B_3$, $r_{3b}(B_1 + B_3)$ . . ., etc.), we must first find the divisors B_3 and $(B_1 + B_3)$. We can obtain B_3 by taking the square root of the sum for the South-East quadrant, viz., $\sqrt{B_3^2}$; and we can obtain $(B_1 + B_3)$ by dividing the grand total of the last two columns by B_3.

To find the factor-saturations for the *group* factor c, we must subtract the 'hierarchy,' formed by multiplying the basic factor-saturations, r_{1b} and r_{2b}, from the North-West quadrant. This leaves a second 'hierarchy,' dependent solely on the group factor c, which can be factorized in the usual way.

(*b*) *Two Group Factors.* Let us now suppose that the South-East quadrant of Table XXII is modified to include a group factor of its own, whose saturation-sum is D_3. The divisors for the row- and column-sums of the North-East quadrant will be B_1 and B_3. The simplest procedure is to determine the ratio B_1/B_3 provisionally (e.g., from the bipolar analysis), calculate C_1 and D_3, and then, if necessary, find improved values for B_1 and B_3 by successive approximation.

(*c*) *Three Group Factors.* Let us take B_1, B_2, B_3 to represent the sums of the basic factor-saturations within the three groups. Then the sums of the non-diagonal submatrices will be B_1B_2, B_1B_3, and B_2B_3. To find B_3, for example, we now calculate $\sqrt{\dfrac{B_1B_3 \,.\, B_2B_3}{B_1B_2}} = \sqrt{B_3^2}$. The divisor for the first column will then be $B_2 + B_3 = \sqrt{\dfrac{B_1B_2 \,.\, B_2B_3}{B_1B_3}} + \sqrt{\dfrac{B_1B_3 \,.\, B_2B_3}{B_1B_2}} = \sqrt{B_2B_3}\left\{\sqrt{\dfrac{B_1B_2}{B_1B_3}} + \sqrt{\dfrac{B_1B_3}{B_1B_2}}\right\}$, which is the formula used on p. 48 above. Similarly for the other divisors.

(*d*) *More than Three Group Factors.* For the general case we simply use the analogue to equation (v), taking the *sums* of the coefficients in the submatrices (obtained by condensation) instead of the individual coefficients themselves.

Method of Least Squares. The foregoing proofs assume that the observed correlations are the *exact* resultants of the basic and group factors. With an empirical table this will not be strictly true. For any given test, however, we can plausibly assume that the *averages* of its observed and of its true correlations are virtually the same. For example, taking test 3 in Table XXII below, we can assume that the average correlation $\bar{r}_{3k}'$ (where the dash denotes an *observed* value) corresponds sufficiently closely with the average of the 'true' correlations ($\bar{r}_{3k}$). Now, with the 'true' values $n\,\bar{r}_{3k} = \Sigma r_{3k} = r_{3b}(B_1 + B_3)$ exactly. If, therefore, we assume $\bar{r}_{3k}' = \bar{r}_{3k}$, we may take $\Sigma r_{3k}' = r_{3b}\,(B_1 + B_3)$. The use of the above formulæ is therefore tantamount to supposing that the discrepancies between the true and the observed values for the several correlations (i.e., the residuals left after all the factors have been extracted) are the result of chance fluctuations, and will therefore sum to zero within the columns or rows (or sections of them) over which the summation extends, provided a sufficient number of tests has been used.

Since, in most psychological investigations, the number of tests on which the factors are based is decidedly small (particularly in the case of the group factors), this assumption is only very approximately true. Hence for any rigorous determination it would be better to proceed by seeking

[1] For the multiplication of partitioned matrices see Cullis, W. E. (1918), *Matrices and Determinoids*, II, pp. 6f., or Aitken, A. C. (1939), *Determinants and Matrices*, pp. 21–25.

TABLE XXII. CONSTRUCTION OF CORRELATION TABLE FOR ONE GROUP FACTOR

Test	1	2	Sum	3	4	Sum	Total
1	$r_{1b}^2 + r_{1c}^2$	$r_{1b}r_{2b} + r_{1c}r_{2c}$	$r_{1b}B_1 + r_{1c}C_1$	$r_{1b}r_{3b}$	$r_{1b}r_{4b}$	$r_{1b}B_3$	—
2	$r_{2b}r_{1b} + r_{2c}r_{1c}$	$r_{2b}^2 + r_{2c}^2$	$r_{2b}B_1 + r_{2c}C_1$	$r_{2b}r_{3b}$	$r_{2b}r_{4b}$	$r_{2b}B_3$	—
Sum	$r_{1b}B_1 + r_{1c}C_1$	$r_{2b}B_1 + r_{2c}C_1$	$B_1^2 + C_1^2$	$r_{3b}B_1$	$r_{4b}B_1$	B_1B_3	—
3	$r_{3b}r_{1b}$	$r_{3b}r_{2b}$	$r_{3b}B_1$	r_{3b}^2	$r_{3b}r_{4b}$	$r_{3b}B_3$	$r_{3b}(B_1 + B_3)$
4	$r_{4b}r_{1b}$	$r_{4b}r_{2b}$	$r_{4b}B_1$	$r_{4b}r_{3b}$	r_{4b}^2	$r_{4b}B_3$	$r_{4b}(B_1 + B_3)$
Sum	$r_{1b}B_3$	$r_{2b}B_3$	B_3B_1	$r_{3b}B_3$	$r_{4b}B_3$	B_3^2	—
Total	—	—	—	$r_{3b}(B_1 + B_3)$	$r_{4b}(B_1 + B_3)$	—	$B_3(B_1 + B_3)$

those particular values for the saturations which, with any postulated factor-pattern,[1] will maximize the likelihood of the test-measurements or test-correlations actually observed. This leads to a procedure which minimizes the square-sums of the residuals, instead of assuming that their algebraic sum will be zero. In practice, however, it is usually found that the differences in the results obtained by the two methods are not large enough to affect any demonstration of the existence or essential nature of the factors so discovered.[2]

II. *Calculation of Rotation Matrix*

Let B be the summed bipolar-matrix, G the summed group factor matrix, and T the required rotation matrix. Then $BT = G$, and $GT' = B$. Writing the last equation explicitly and substituting the known values for G and B from Table XVIII A and C, we have

$$\begin{bmatrix} \cdot8816 & \cdot0579 & \cdot0000 & \cdot0000 \\ 1\cdot7389 & \cdot6299 & \cdot0000 & \cdot0000 \\ 1\cdot5875 & \cdot0000 & \cdot8244 & \cdot0000 \\ 1\cdot7134 & \cdot0000 & \cdot0000 & 1\cdot4228 \end{bmatrix} \begin{bmatrix} a & e & j & p \\ b & f & k & q \\ c & g & m & r \\ d & h & n & s \end{bmatrix} = \begin{bmatrix} \cdot8506 & \cdot1568 & \cdot1062 & \cdot1453 \\ 1\cdot7352 & \cdot4177 & \cdot4628 & -\cdot1453 \\ 1\cdot6300 & \cdot4680 & -\cdot5690 & \cdot0000 \\ 1\cdot9681 & -1\cdot0425 & \cdot0000 & \cdot0000 \end{bmatrix}$$

where lower case letters are substituted for the several columns of T'. The above matrix equation provides us with four systems of simultaneous linear equations, each system containing four equations to be solved for four unknown quantities. Let us take the first column of unknown quantities, namely, a, b, c, and d. Changing to the ordinary notation, the matrix equation for these quantities may be expanded as follows :

$$\begin{aligned} \cdot8816a + \cdot0579b &= \cdot8506 \quad . \quad . \quad . \quad \text{(i)} \\ 1\cdot7389a + \cdot6299b &= 1\cdot7352 \quad . \quad . \quad . \quad \text{(ii)} \\ 1\cdot5875a + \cdot8244c &= 1\cdot6300 \quad . \quad . \quad . \quad \text{(iii)} \\ 1\cdot7134a + 1\cdot4228d &= 1\cdot9681 \quad . \quad . \quad . \quad \text{(iv)} \end{aligned}$$

Multiplying (i) by $\cdot6299/\cdot0579 = 10\cdot8791$, we have :

$$9\cdot5910a + \cdot6299b = 9\cdot2538 \quad . \quad . \quad . \quad \text{(v)}$$

Subtracting (ii) from (v) we get :

$$7\cdot8521a = 7\cdot5186.$$

Hence $a = \cdot9575.$

Substituting for a, we obtain

from (ii) $1\cdot6651 + \cdot6299b = 1\cdot7352$. Hence $b = \cdot1113$.
from (iii) $1\cdot5200 + \cdot8244c = 1\cdot6300$. Hence $c = \cdot1334$.
from (iv) $1\cdot6405 + 1\cdot4228d = 1\cdot9681$. Hence $d = \cdot2303$.

The figures thus obtained form the first row of T in Table XIX A. We can solve the equations for the three remaining columns of T' in the same way. It will be noted that the multiplier for the first equation (10·8791) is the same for all four columns. Hence in practice it is quicker to calculate corresponding steps for the four columns at the same time.

When all the values for the rotation matrix have been calculated by this method, the figures should be checked by verifying that the square-sum for each of the four rows amounts to 1·0000, and the product-sum for every pair of rows to zero.

I should like to express my indebtedness to my colleagues, Dr. Banks and Mr. Summerfield, who have been good enough to read this article in proof, and check many of the calculations.

[1] I here use the phrase ' factor-pattern ' to denote the general scheme of positive, negative, or zero saturations, regardless of their numerical value. The full procedure would require us to calculate the amounts of the maximum likelihoods obtainable with *different* hypothetical factor-patterns (e.g., a two-factor pattern, a general-plus-bipolar factor-pattern, a basic-plus-group factor-pattern, and a ' simple structure ' of group factors only), and then adopt, as provisionally proved, that hypothesis which furnished the largest value for its maximum likelihood—a proof by ' maximal maximum.'

[2] More detailed proofs for these various cases are given in the roneo'd *Laboratory Notes on Factor Analysis : Group Factor Method.* For a fuller discussion of the initial definitions and postulates see 7, pp. 246f., which incorporates the essentials from earlier memoranda.

FACTOR ANALYSIS BY MAXIMUM LIKELIHOOD: A CORRECTION

By D. N. LAWLEY
The Mathematical Institute, University of Edinburgh

In a recent paper (1) Mr. W. G. Emmett has quoted formulæ for the standard errors of factor loadings estimated by my method of maximum likelihood. The use of the formulæ, however, requires an additional qualification, since the derivation given in the paper (2) from which they were quoted is itself in need of correction.

It has now been found that the estimated loadings obtained for a given set of tests in one factor are slightly correlated with estimated loadings obtained for the same set of tests in another factor. This in turn means that the standard errors given by the formula for the estimated loadings in the rth factor are correct only if the values of loadings in previous factors are regarded as already given.

I should add that this correction in no way detracts from the value of Mr. Emmett's own paper.

REFERENCES

1. Emmett, W. G. (1949). 'Factor Analysis by Lawley's Method of Maximum Likelihood. *Brit. J. Psych. Stat. Sect.*, 90–97.
2. Lawley, D. N. (1949). 'Problems in Factor Analysis.' *Proc. Roy. Soc. Edin.* *A*, LXII, 1–6.